

CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one member,  
 CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
 CC degree of hybridisation to both classes is determined from the label on  
 CC the amplicon. From the ratio of labels hybridised to the two classes of  
 CC oligomers, the degree of methylation is calculated. The method is used:  
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 CC and of a wide range of diseases, e.g. cancer, disorders of the central  
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 CC particularly by detecting mutations or single nucleotide polymorphisms  
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
 CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention  
 CC  
 XX Sequence 530 BP; 172 A; 230 C; 76 G; 52 T; 0 U; 0 Other;  
 SQ

Query Match 65.4%; Score 17; DB 6; Length 530;  
 Best Local Similarity 80.0%; Pred. No. 2.7e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGATCTACGAA 26  
 DB 443 CCACGGCGTCGACGATCTACGAA 467

RESULT 29  
 ABQ14534/C  
 ID ABQ14534 standard; DNA; 530 BP.  
 XX  
 AC ABQ14534;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 1125.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.

XX Homo sapiens.  
 OS  
 XX WO200218632-A2.  
 FN  
 XX PD 07-MAR-2002.  
 XX  
 XX 01-SEP-2001; 2001WO-EP010074.  
 XX  
 XX 01-SEP-2000; 2000DE-01043826.  
 PR  
 XX 05-SEP-2000; 2000DE-01044543.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 PA  
 XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
 PI  
 XX WPI; 2002-371829/40.  
 DR

PT Determining the degree of cytosine methylation in genomic DNA, useful for  
 PT diagnosis and prognosis, comprises selective hybridization of amplicons  
 PT from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
 XX  
 XX This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one member,  
 CC

CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
 CC degree of hybridisation to both classes is determined from the label on  
 CC the amplicon. From the ratio of labels hybridised to the two classes of  
 CC oligomers, the degree of methylation is calculated. The method is used:  
 CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
 CC and of a wide range of diseases, e.g. cancer, disorders of the central  
 CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
 CC particularly by detecting mutations or single nucleotide polymorphisms  
 CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
 CC investigating cell differentiation. The method allows the methylation  
 CC status of many C residues to be determined simultaneously. ABQ13410-  
 CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
 CC for determining the degree of cytosine methylation described in the  
 CC disclosure of the invention  
 CC

XX Sequence 530 BP; 52 A; 76 C; 230 G; 172 T; 0 U; 0 Other;  
 SQ

Query Match 65.4%; Score 17; DB 6; Length 530;  
 Best Local Similarity 80.0%; Pred. No. 2.7e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGATCTACGAA 26  
 DB 88 CCACGGCGTCGACGATCTACGAA 64

RESULT 30  
 AAA55205/C  
 ID AAA55205 standard; DNA; 879 BP.

XX  
 AC AAA55205;  
 XX  
 DT 30-AUG-2000 (first entry)  
 XX  
 DE C. symbiosum open reading frame nucleotide sequence SEQ ID NO:37.

XX Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;  
 XX characterisation; archae; therapeutic; industrial; laboratory; ds.  
 KW  
 XX Cenarchaeum symbiosum.

OS  
 XX WO200018909-A2.  
 FN  
 XX PD 06-APR-2000.

XX 29-SEP-1999; 99WO-US022752.  
 XX  
 XX 29-SEP-1998; 98US-0102294P.  
 PR  
 XX (DIVE-) DIVERSA CORP.

XX Swanson RV, Feldman RA, Schleper C;  
 PI  
 XX WPI; 2000-293148/25.  
 DR  
 XX P-PSDB; AAY90930.

XX New nucleic acids and proteins isolated from the non-thermophilic  
 PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the  
 PT physiology of these archae and in therapeutic, industrial or laboratory  
 PT techniques.  
 PT  
 XX Claim 7; Page 149-151; 210pp; English.

XX AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids and  
 CC proteins isolated from the non-thermophilic crenarchaeote Cenarchaeum  
 CC symbiosum. The nucleic acids and proteins identified in the present  
 CC invention are useful in characterising the physiology of these archae and  
 CC can be used in therapeutic, industrial or laboratory techniques. AAA55227  
 CC to AAA55260 represent promoter sequences from Cenarchaeum symbiosum.  
 CC AAA55261 to AAA55269 represent PCR primers and probes used in examples  
 CC from the present invention  
 CC

XX Sequence 879 BP; 142 A; 270 C; 279 G; 188 T; 0 U; 0 Other;  
 SQ

Query Match 65.4%; Score 17; DB 3; Length 879;  
 Best Local Similarity 80.0%; Pred. NO. 2.8e+02;  
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 ACAACGGCGAGCGCCGGAATCTACGA 25  
 Db 434 ACCTCGAGAGCGCCGAGTCTACAA 410

Search completed: June 20, 2004, 10:17:36  
 Job time : 165.522 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1214.98 Seconds  
(without alignments)  
639.034 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26

Sequence: 1 acaacggcaggccgaatctacgaa 26

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estl:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estom:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vit:\*

21: em\_gss\_fun:\*

22: em\_gss\_mus:\*

23: em\_gss\_pro:\*

24: em\_gss\_rod:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.8	80.0	298	CC930679	ZMWBBC054
C 2	19.6	75.4	570	CD667796	NCEST3B87
C 3	19.6	75.4	1025	AG086420	Pan trogl
C 4	19.2	73.8	648	BI354861	GM28636.5

C 78	17.6	67.7	985	29	CNS01RER	AL157315 Anopheles	C 151	17	65.4	574	14	CFI03480
C 79	17.4	66.9	769	10	BF141737	BF141737 601787601	C 152	17	65.4	578	14	CB404010
C 80	17.4	66.9	1100	12	B1211982	B1211982 602936449	C 153	17	65.4	584	14	CD666771
C 81	17.2	66.2	190	29	CG907394	CG907394 ZMWB052	C 154	17	65.4	588	13	BU571686
C 82	17.2	66.2	267	14	CA742056	CA742056 wetic.pk0	C 155	17	65.4	595	14	CD665628
C 83	17.2	66.2	325	9	AV836085	AV836085 AV836085	C 156	17	65.4	596	12	BM086115
C 84	17.2	66.2	371	9	AV918494	AV918494 AV918494	C 157	17	65.4	597	14	CA179494
C 85	17.2	66.2	377	13	BG418303	BG418303 HVSMEK002	C 158	17	65.4	612	28	BH210205
C 86	17.2	66.2	390	13	CA009415	CA009415 HU1A24r	C 159	17	65.4	617	10	BH210205
C 87	17.2	66.2	431	9	AJ343400	AJ343400 AJ343400	C 160	17	65.4	623	14	CF863327
C 88	17.2	66.2	433	13	BG768842	BG768842 EBR008_SQ	C 161	17	65.4	625	12	BI238607
C 89	17.2	66.2	460	14	CA631465	CA631465 wleln.pk0	C 162	17	65.4	625	14	CA405369
C 90	17.2	66.2	496	12	BN098164	BN098164 ESP103_SQ	C 163	17	65.4	626	12	BI238566
C 91	17.2	66.2	500	9	EU191490	EU191490 AU191490	C 164	17	65.4	627	29	CG602473
C 92	17.2	66.2	508	13	CA006074	CA006074 HU05114u	C 165	17	65.4	634	12	BI163007
C 93	17.2	66.2	518	14	CB354137	CB354137 ZF001-P00	C 166	17	65.4	641	14	CD435659
C 94	17.2	66.2	586	14	CB877243	CB877243 HC07C05Y	C 167	17	65.4	642	14	CF611457
C 95	17.2	66.2	641	14	CD860447	CD860447 TE-002H11	C 168	17	65.4	649	9	AI513469
C 96	17.2	66.2	641	14	CD860447	CD860447 TE-002H11	C 169	17	65.4	656	10	BE188342
C 97	17.2	66.2	699	12	BJ468153	BJ468153 BJ468153	C 170	17	65.4	656	12	BI366362
C 98	17.2	66.2	702	28	CC071842	CC071842 CSU-K33r	C 171	17	65.4	659	14	CA224539
C 99	17.2	66.2	724	13	BX073676	BX073676 BX073676	C 172	17	65.4	660	28	BZ892674
C 100	17.2	66.2	750	28	BH987721	BH987721 Oei14A06	C 173	17	65.4	667	14	CF864061
C 101	17.2	66.2	759	29	CNS038G9	AL240354 Tetraodon	C 174	17	65.4	671	12	BI228894
C 102	17.2	66.2	764	14	CA358950	CA358950 631799 NC	C 175	17	65.4	691	29	CD819340
C 103	17.2	66.2	841	9	AI069114	AI069114 mgae0005C	C 176	17	65.4	703	14	CD861967
C 104	17.2	66.2	861	28	AQ325329	AQ325329 mgxb0022L	C 177	17	65.4	715	14	CD826857
C 105	17.2	66.2	969	12	BG259769	BG259769 602378029	C 178	17	65.4	715	14	CF393044
C 106	17.2	66.2	981	10	BZ739529	BZ739529 601556465	C 179	17	65.4	721	29	C5633838
C 107	17.2	66.2	1009	29	CG935282	CG935282 ZMWB0054	C 180	17	65.4	734	13	BQ804860
C 108	17.2	66.2	1027	10	BF216448	BF216448 601884507	C 181	17	65.4	747	28	BH513243
C 109	17.2	66.2	118	10	BF116130	BF116130 7n77n11.X	C 182	17	65.4	756	9	AA979545
C 110	17.2	66.2	163	29	CG708761	CG708761 OGWDK76TH	C 183	17	65.4	766	9	AA950536
C 111	17.2	66.2	166	28	AZ821274	AZ821274 2M0094A03	C 184	17	65.4	778	12	BI114255
C 112	17.2	66.2	229	28	AQ962931	AQ962931 LERGI37TR	C 185	17	65.4	788	29	BX163589
C 113	17.2	66.2	240	9	AI188735	AI188735 AV188735	C 186	17	65.4	794	28	BZ734543
C 114	17.2	66.2	291	12	BM158980	BM158980 NXIV_042	C 187	17	65.4	809	29	CG705111
C 115	17.2	66.2	331	29	CG795376	CG795376 SALK_0803	C 188	17	65.4	814	29	BX148091
C 116	17.2	66.2	340	14	CF845658	CF845658 PSHB034XA	C 189	17	65.4	815	12	BG684546
C 117	17.2	66.2	343	12	BI486572	BI486572 RH70020.5	C 190	17	65.4	816	28	BZ54434
C 118	17.2	66.2	348	14	CB827912	CB827912 LJNEST80F	C 191	17	65.4	820	29	CG938758
C 119	17.2	66.2	349	14	CB031405	CB031405 TGEStzyd2	C 192	17	65.4	827	29	CG044322
C 120	17.2	66.2	360	9	AV192051	AV192051 AV192051	C 193	17	65.4	863	29	CG044322
C 121	17.2	66.2	376	10	BH770031	BH770031 IL3-ET011	C 194	17	65.4	885	28	CG24943
C 122	17.2	66.2	383	12	BG159027	BG159027 RHIZ2_41	C 195	17	65.4	904	12	BI106690
C 123	17.2	66.2	399	12	BG274127	BG274127 WHE2231_H	C 196	17	65.4	923	29	CG679436
C 124	17.2	66.2	399	14	CB384983	CB384983 OSTF015D1	C 197	17	65.4	940	28	CG239746
C 125	17.2	66.2	399	14	CB403971	CB403971 OSTF015D1	C 198	17	65.4	967	12	BG251378
C 126	17.2	66.2	400	12	BG413258	BG413258 ETESted19	C 199	17	65.4	969	29	CNS02MJK
C 127	17.2	66.2	427	9	AU282923	AU282923 AU282923	C 200	17	65.4	1008	13	BG672812
C 128	17.2	66.2	441	14	CA658408	CA658408 w1m0.pk04	C 201	17	65.4	1058	28	BZ577858
C 129	17.2	66.2	447	10	BE437034	BE437034 WHE0562_F	C 202	17	65.4	1122	29	CNS05Q2Y
C 130	17.2	66.2	455	28	AZ132296	AZ132296 OSUNBD006	C 203	17	65.4	1153	12	BZ552819
C 131	17.2	66.2	464	28	AQ535924	AQ535924 RFI-11-4	C 204	17	65.4	1153	12	BZ552819
C 132	17.2	66.2	473	28	CA109661	CA109661 CIT-HSP-2	C 205	17	65.4	1104	10	BE738725
C 133	17.2	66.2	479	14	CB405915	CB405915 OSTF059E1	C 206	16.8	64.6	104	13	BU896955
C 134	17.2	66.2	492	29	CG820351	CG820351 100001A22	C 207	16.8	64.6	385	12	BM147165
C 135	17.2	66.2	507	28	AQ879656	AQ879656 HS_4821_A	C 208	16.8	64.6	439	13	BY447416
C 136	17.2	66.2	507	28	AQ879706	AQ879706 HS_4821_A	C 209	16.8	64.6	582	10	BF252054
C 137	17.2	66.2	510	12	EG516742	EG516742 ETESted61	C 210	16.8	64.6	649	12	BG787388
C 138	17.2	66.2	516	12	BG904715	BG904715 TAlr1134D	C 211	16.8	64.6	663	12	BI669703
C 139	17.2	66.2	517	13	CA001330	CA001330 HS18-06u	C 212	16.8	64.6	993	29	CNS05M64
C 140	17.2	66.2	521	29	AG224590	AG224590 Lotus cor	C 213	16.8	64.6	1060	12	BM924514
C 141	17.2	66.2	534	14	CF753771	CF753771 EST-77-2-	C 214	16.6	63.8	145	10	BF116847
C 142	17.2	66.2	535	9	AI996716	AI996716 701668055	C 215	16.6	63.8	174	14	CF556237
C 143	17.2	66.2	540	28	BZ895417	BZ895417 NARP0.01	C 216	16.6	63.8	179	12	BG487500
C 144	17.2	66.2	547	28	BH643435	BH643435 1008057F0	C 217	16.6	63.8	234	14	CF865821
C 145	17.2	66.2	556	13	CA005431	CA005431 HU10703u	C 218	16.6	63.8	236	14	CA647697
C 146	17.2	66.2	564	9	AI456544	AI456544 LD36595.5	C 219	16.6	63.8	243	13	BQ700211
C 147	17.2	66.2	565	9	AI770273	AI770273 SAL3_E02	C 220	16.6	63.8	274	29	CG712165
C 148	17.2	66.2	568	14	CF882197	CF882197 tric029x1	C 221	16.6	63.8	307	9	AU070159
C 149	17.2	66.2	570	9	AI770270	AI770270 SAL3_D04	C 222	16.6	63.8	315	14	BQ760612
C 150	17.2	66.2	574	12	BU159696	BU159696 BU159696	C 223	16.6	63.8	345	14	CA633072

```
224 16.6 63.8 355 9 AA745637
c 225 16.6 63.8 357 14 CA651760
226 16.6 63.8 360 14 CB867769
227 16.6 63.8 361 9 AA166277
c 228 16.6 63.8 377 14 CR81912
229 16.6 63.8 377 14 CR86362
230 16.6 63.8 385 14 CA603922
231 16.6 63.8 392 14 CF556236
232 16.6 63.8 396 12 BJ482339
c 233 16.6 63.8 397 29 CC951609
c 234 16.6 63.8 416 13 BQ660741
c 235 16.6 63.8 421 14 CD937856
c 236 16.8 63.8 424 14 CD882523
c 237 16.6 63.8 427 10 BE591405
238 16.6 63.8 434 14 CD57595
239 16.6 63.8 444 9 AU183310
240 16.6 63.8 444 9 CB866388
241 16.6 63.8 445 14 CB867017
242 16.6 63.8 446 10 BE018744
243 16.6 63.8 450 9 AL817225
c 244 16.6 63.8 455 14 CB213437
245 16.6 63.8 459 28 BZ334215
c 246 16.6 63.8 463 13 BQ658174
c 247 16.6 63.8 466 28 AZ867494
c 248 16.6 63.8 467 9 AL505858
249 16.6 63.8 480 14 CF643680
250 16.6 63.8 502 14 CA180087
251 16.6 63.8 505 14 CF564588
252 16.6 63.8 512 12 BG909982
c 253 16.6 63.8 514 14 CF564589
c 254 16.6 63.8 527 29 CNG03HIW
c 255 16.6 63.8 528 9 AV632220
256 16.6 63.8 528 14 CF557920
257 16.6 63.8 529 14 CA642629
258 16.6 63.8 531 14 CB866928
259 16.6 63.8 537 12 BG605539
c 260 16.6 63.8 537 14 CF797231
c 261 16.6 63.8 538 14 CF274843
c 262 16.6 63.8 539 14 CA183146
c 263 16.6 63.8 549 14 CA637222
c 264 16.6 63.8 550 14 CD883343
265 16.6 63.8 551 13 BQ800717
266 16.6 63.8 552 13 BQ800717
267 16.6 63.8 553 14 CB866423
268 16.6 63.8 558 14 CA596282
269 16.6 63.8 559 14 CA700439
270 16.6 63.8 562 14 CA221192
271 16.6 63.8 563 29 CC967754
272 16.6 63.8 565 12 BG632274
273 16.6 63.8 566 13 BQ458375
274 16.6 63.8 568 13 CA112709
275 16.6 63.8 570 14 CK131140
276 16.6 63.8 572 14 CF555545
277 16.6 63.8 577 14 CF797396
c 278 16.6 63.8 577 14 CF942440
279 16.6 63.8 578 28 AQ619676
280 16.6 63.8 579 14 CA239443
281 16.6 63.8 580 12 BG487499
282 16.6 63.8 582 9 AV397650
283 16.6 63.8 585 12 BG606523
c 284 16.6 63.8 587 14 CF560327
c 285 16.6 63.8 587 14 CF565633
286 16.6 63.8 587 14 CF565634
c 287 16.6 63.8 590 29 CG209669
c 288 16.6 63.8 595 13 BU999353
289 16.6 63.8 598 14 CA733694
c 290 16.6 63.8 598 14 CF565484
291 16.6 63.8 599 28 CG146982
c 292 16.6 63.8 600 28 BZ895461
c 293 16.6 63.8 606 14 CA728823
c 294 16.6 63.8 608 14 CF565483
c 295 16.6 63.8 617 14 CF560045
296 16.6 63.8 622 14 CA606145

AA745637 ny69f10.s
CA651760 wreln.pk1
CB867769 HCO1C24w
AA166277 ms48d11.r
CF81912 trlc085xn
CF86362 trlc030xa
CA603922 wr1.pk003
CF556236 111501SH1
BJ482339 BJ482339
CC951609 BOHC21TR
BQ660741 HI05D16U
CD937856 OV.108B21
CD882523 FL.110C04
BE591405 WHE1659-1
CD57595 HO15J18S
AU183310 AU183310
CB866388 HCO7E20w
CB867017 HCO5F03w
BE018744 bb84d07.Y
AL817225 AL817225
CB213437 OML03717
BZ334215 bx77f05.9
BQ658174 HA05G18u
AZ867494 2M0178520
AL505858 AL505858
CF643680 K11.D09.F
CA180087 SCCST300
CF564588 1115092H0
BG909982 Talr1112E
CF564589 1115092H0
AL244337 Tetraodon
AV632220 AV632220
CF557920 1115032F0
CA642629 wreln.pk0
CB866928 HCO5J17w
BG605539 WHE2332.H
CF797231 NCBST3C56
CF274843 NCBST3D34
CA183146 SCMCST316
CA637222 wre1.pk00
CD883343 FL.112P21
CF560047 1115052H0
BQ800717 san37c01.
CB866423 HCO7D01w
CA596282 wpa1c.pk0
CA700439 wkmlc.pk0
CA221192 SCBGPLA05
CC967754 BO1FA54TF
BG632274 GR06015.3
BQ458375 HA05G18r
CA112709 SCEQUB106
CK131140 GH05220.3
CF555545 1115009C0
CF797396 NCSTP3C57
CF942440 NCSTeqab9
AQ619676 HS.5185.A
CA239443 SCRFPL503
BG487499 EM1.65.G1
AV397650 AV397650
BG606523 WHE2957.C
CF560327 1115055E0
CF565633 1115101G0
CF565634 1115101G0
CG209669 OGVDU59TH
BU999353 HI14D22r
CA733694 wlp1c.pk0
CF565484 1115100F0
CG146982 PU1EJ32PB
BZ895461 NARP1.01
CA728823 wd1c.pk0
CF565483 1115100F0
CF560045 1115052H0
CA606145 wr1.pk006

297 15.6 63.8 624 14 CA187672
c 298 15.6 63.8 624 14 CF563749
299 15.6 63.8 624 14 CF563751
300 15.6 63.8 624 14 CF565485

ALIGNMENTS

CC930679 298 bp DNA linear GSS 11-AUG-2003
ZMMBBC0541G09r ZMMBbc Zea mays subsp. mays genomic clone
ZMMBBC0541G09 3', genomic survey sequence.
CC930679
CC930679.1 GI:33581777
GSS.
Zea mays subsp. mays (maize)
Zea mays subsp. mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 298)
Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Rouzar,K., Fuks,G., Yu,Y., Wing,R. and Messing,J.
Sequencing of the maize genome at PGR (2003b)
Unpublished (2003)
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: SP6
Class: BAC ends
High quality sequence start: 124.
Location/Qualifiers
1..298
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMMBBC0541G09"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBbc"
/note="Vector: pTARBAC1.3; Site_1: BamHI; Site_2: BamHI"

Query Match 80.0%; Score 20.8; DB 29; Length 298;
Best Local Similarity 91.7%; Pred. No. 3.9e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AACGCGAGGCGCGAATCTACGAA 26
|||||
Db 181 AACGCGAGGCGCGAATCTCGAA 158

RESULT 2
CD667796 570 bp mRNA linear EST 23-JUN-2003
LOCUS NCST3B97e08.y1 Nc-1 Tachyzoite cDNA Library Neospora caninum cDNA
5', mRNA sequence.
DEFINITION
CD667796
ACCESSION CD667796.1 GI:32162466
VERSION EST.
KEYWORDS Neospora caninum
ORGANISM Neospora caninum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Neospora.
1 (bases 1 to 570)
```



XhoI; Sized fractionated cDNAs were directly ligated into pOT2."

## ORIGIN

Query Match 73.8%; Score 19.2; DB 12; Length 648;  
Best Local Similarity 84.0%; Pred. No. 2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 ACAACGGCGAGCGCGGAATCTAGCA 25  
|||||  
Db 577 AGAAGCGGACGCGCGGAATCTGCCA 601

## RESULT 5

BZ567119 875 bp DNA linear GSS 17-DEC-2002  
LOCUS pacs2-164 6781.x1 pacs2-164 Pseudomonas aeruginosa genomic clone  
DEFINITION pacs2-164 6781, genomic survey sequence.

ACCESSION BZ567119

VERSION BZ567119.1 GI:27197852

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

REFERENCE 1 (bases 1 to 875)  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

AUTHORS Spencer, D.H., Raymond C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.B., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

JOURNAL J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

## FEATURES

source

1..875  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacs2-164 6781"  
/clone\_lib="pacs2-164"  
/note="Clinical isolate 2-164 Whole genomic shotgun library."

## ORIGIN

Query Match 73.8%; Score 19.2; DB 28; Length 875;  
Best Local Similarity 87.5%; Pred. No. 2.2e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 ACAACGGCGAGCGCGGAATCTACG 24  
|||||  
Db 787 ACAACGGCGAGCGCGGAATCTTCG 810

## RESULT 6

BY077170/c 367 bp mRNA linear EST 06-DEC-2002  
LOCUS BY077170 RIKEN full-length enriched, pooled tissues, adult spleen,  
DEFINITION etc. Mus musculus cDNA clone K630008F15 5', mRNA sequence.

ACCESSION BY077170

VERSION BY077170.1 GI:26178644

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 367)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Boro, H., Kondo, S.,  
Nikaido, I., Oeato, N., Saito, R., Suzuki, H., Yamakawa, I.,  
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,  
Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Buit, C.,  
Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H.,  
Batakou, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusci, V.,  
Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,  
Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,  
Gustibaldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S.,  
Kawachi, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,  
Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,  
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,  
Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G.,  
Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S.,  
Ravasi, F., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,  
Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,  
Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,  
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,  
Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,  
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,  
Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,  
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Harai, A., Hashizume, M., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,  
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S.,  
Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

22354683

12465851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,

Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,

Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, K.,

Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,

Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and

Hayashizaki, Y. Direct Submission

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

Location/Qualifiers

1..367

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="K630008F15"

/clone\_lib="RIKEN full-length enriched, pooled tissues,

adult spleen, etc."

## FEATURES

source



USDA-PP-122000-752 Immature Ovaries from field-collected Valencia Sweet Orange (*Citrus sinensis* (L.) Osbeck) *Citrus sinensis* cDNA clone MVF-55\_E04 5', mRNA sequence.

CP507251  
CP507251.1 GI:34522435  
EST.  
Citrus sinensis  
Citrus sinensis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucosids II; Sapindales; Rutaceae; Citrus.

REFERENCE  
1 (bases 1 to 458)  
Chaparro, J.X., Bausher, M.G., Dang, P., Hunter, W.B., McKenzie, C.L., Niede, R.P., and Shatters, R.G. Jr.  
Expressed Sequence Tags from Immature Ovaries of Valencia Sweet Orange (*Citrus sinensis* (L.) Osbeck)  
Unpublished (2003)  
JOURNAL  
COMMENT  
Contact: Chaparro, JX  
Horticulture and Breeding Research Unit  
USDA, ARS, U. S. Horticultural Research Laboratory  
2001, South Rock Road, Fort Pierce, FL 34945, USA  
Tel: 772 462 5830  
Fax: 772 462 5986  
Email: jchaparro@ushrl.ars.usda.gov  
Seq primer: T3 Primer.

FEATURES  
source  
1..458  
/organism="Citrus sinensis"  
/mol\_type="mRNA"  
/cultivar="Valencia"  
/db\_xref="taxon:2711"  
/clone="MVF-55\_E04"  
/tissue\_type="Immature ovaries"  
/clone\_lib="Immature Ovaries from field-collected Valencia Sweet Orange (*Citrus sinensis* (L.) Osbeck)"  
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Standard library construction protocols from Stratagene cDNA synthesis kit (cat.# 200401-5) and Uni-ZAP XR vector kit (cat.# 237211) were followed using poly(A) RNA."

ORIGIN  
Query Match 71.5%; Score 18.6; DB 14; Length 458;  
Best Local Similarity 84.0%; Pred. No. 3.2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26  
DB 197 CAACGGCGAGCGCCGAATCTACGAA 173

RESULT 10  
AI5333801/c  
AI5333801/c  
LOCUS  
DEFINITION  
SD05764.3prime SD Drosophila melanogaster Schneider L2 cell culture  
POT2 Drosophila melanogaster cDNA clone SD05764 3prime, mRNA  
sequence.

ACCESSION  
AI533801  
VERSION  
AI533801.1 GI:4447936  
KEYWORDS  
EST.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 510)  
Harvey, D., Brokstein, P., Hong, L., Evans-Hoim, M., Su, C., Tsang, G., Lewis, S., and Rubin, G.M.  
BDGP/FHMI Drosophila EST Project  
Unpublished (2001)  
JOURNAL  
COMMENT  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab

USDA-PP-122000-752 Immature Ovaries from field-collected Valencia Sweet Orange (*Citrus sinensis* (L.) Osbeck) *Citrus sinensis* cDNA clone MVF-55\_E04 5', mRNA sequence.

CP507251  
CP507251.1 GI:34522435  
EST.  
Citrus sinensis  
Citrus sinensis  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eucosids II; Sapindales; Rutaceae; Citrus.

REFERENCE  
1 (bases 1 to 458)  
Chaparro, J.X., Bausher, M.G., Dang, P., Hunter, W.B., McKenzie, C.L., Niede, R.P., and Shatters, R.G. Jr.  
Expressed Sequence Tags from Immature Ovaries of Valencia Sweet Orange (*Citrus sinensis* (L.) Osbeck)  
Unpublished (2003)  
JOURNAL  
COMMENT  
Contact: Chaparro, JX  
Horticulture and Breeding Research Unit  
USDA, ARS, U. S. Horticultural Research Laboratory  
2001, South Rock Road, Fort Pierce, FL 34945, USA  
Tel: 772 462 5830  
Fax: 772 462 5986  
Email: jchaparro@ushrl.ars.usda.gov  
Seq primer: T3 Primer.

FEATURES  
source  
1..458  
/organism="Citrus sinensis"  
/mol\_type="mRNA"  
/cultivar="Valencia"  
/db\_xref="taxon:2711"  
/clone="MVF-55\_E04"  
/tissue\_type="Immature ovaries"  
/clone\_lib="Immature Ovaries from field-collected Valencia Sweet Orange (*Citrus sinensis* (L.) Osbeck)"  
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Standard library construction protocols from Stratagene cDNA synthesis kit (cat.# 200401-5) and Uni-ZAP XR vector kit (cat.# 237211) were followed using poly(A) RNA."

ORIGIN  
Query Match 71.5%; Score 18.6; DB 14; Length 458;  
Best Local Similarity 84.0%; Pred. No. 3.2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26  
DB 197 CAACGGCGAGCGCCGAATCTACGAA 173

RESULT 10  
AI5333801/c  
AI5333801/c  
LOCUS  
DEFINITION  
SD05764.3prime SD Drosophila melanogaster Schneider L2 cell culture  
POT2 Drosophila melanogaster cDNA clone SD05764 3prime, mRNA  
sequence.

ACCESSION  
AI533801  
VERSION  
AI533801.1 GI:4447936  
KEYWORDS  
EST.  
SOURCE  
Drosophila melanogaster (fruit fly)  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 510)  
Harvey, D., Brokstein, P., Hong, L., Evans-Hoim, M., Su, C., Tsang, G., Lewis, S., and Rubin, G.M.  
BDGP/FHMI Drosophila EST Project  
Unpublished (2001)  
JOURNAL  
COMMENT  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: [http://www.fruitfly.org/EST\\_estfruitfly.berkeley.edu](http://www.fruitfly.org/EST_estfruitfly.berkeley.edu)  
Based upon the presence of a XhoI site followed by a run of 14 or more T residues at the beginning of the sequence, this clone probably contains an inverted insert. The resulting Poly-T sequence has been removed.  
Plate: 57 row: F column: 4  
High quality sequence stop: 403.  
Location/Qualifiers  
1..510  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="SD05764"  
/lab\_host="DH5-alpha"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell culture pOT2"  
/note="Vector: pOT2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2. Plasmid cDNA library."

ORIGIN  
Query Match 71.5%; Score 18.6; DB 9; Length 510;  
Best Local Similarity 84.0%; Pred. No. 3.3e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGAATCTACGA 25  
DB 191 AGAACGGCGAGCGCCGAATCTCCGA 167

RESULT 11  
BH85898/c  
BH85898/c  
LOCUS  
DEFINITION  
LB00469a.d.SP6.1 Leishmania major Friedlin BAC Library Leishmania major genomic clone LB00469a, Genomic survey sequence.

ACCESSION  
BH85898  
VERSION  
BH85898.1 GI:22130293  
KEYWORDS  
GSS  
SOURCE  
Leishmania major  
Leishmania major  
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

REFERENCE  
1 (bases 1 to 519)  
Myler, P.J., Vogt, C., Munden, H., Robertson, L., Sisk, E., Fazalinia, G., Aggarwal, G., Nelson, S., Seyler, A., Worthey, E., Stuart, K., and Ragland, M.  
Leishmania major Friedlin BAC End Sequences  
Unpublished (2002)  
Other GSSs: LB00469a.d.T7.1  
Contact: Myler PJ  
Seattle Biomedical Research Institute  
4 Nickerson Street, Seattle, WA 98109-1651, USA  
Tel: 206 284-8846  
Fax: 206 284-0313  
Email: [mylerpj@sbri.org](mailto:mylerpj@sbri.org)  
Seq primer: SP6  
Class: BAC ends.

FEATURES  
source  
1..519  
/organism="Leishmania major"  
/mol\_type="genomic DNA"  
/strain="Friedlin"  
/db\_xref="taxon:5664"  
/clone="LB00469a"  
/lab\_host="E. coli GeneHogs + TrfA"  
/clone\_lib="Leishmania major Friedlin BAC Library"  
/note="Vector: pCG270; Site 1: HindIII; Genomic DNA from Leishmania major Friedlin in agarose blocks was partially digested with HindIII, size selected, and ligated with HindIII-digested pCG270 vector DNA. 10368 clones were picked and arrayed in 384- and 96-well plates. Library



construction and arraying was carried out by ResGen Corporation and clones and filters are available from them"

## ORIGIN

Query Match 71.5%; Score 18.6; DB 28; Length 519;  
Best Local Similarity 84.0%; Pred. No. 3.3e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26

Db 378 CACGGCGAGCGCGCGAATCAACGAA 354

## RESULT 12

BQ623129/c

LOCUS

DEFINITION USDA-FP 00220 Ridge pineapple sweet orange entire seedling Citrus sinensis cDNA clone USDA-FP\_00220 5', mRNA sequence.

ACCESSION BQ623129

VERSION BQ623129.1 GI:21650298

KEYWORDS EST.

SOURCE Citrus sinensis

ORGANISM Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Sapindales; Rutaceae; Citrus.

## REFERENCE

AUTHORS

1 (bases 1 to 554) Bausher, M., McKendree, W., Dang, P., Chaparro, J., Shatters, R., Hunter, W. and Niedz, R.

Expressed sequence tags isolated from entire sweet orange (C. sinensis L. Osbeck) seedling

UNPUBLISHED (2003)

CONTACT: Michael Bausher

US Horticultural Research

USDA - ARS

2001 South Rock Rd., Fort Pierce, FL 34945, USA

Tel: (772) 462-5918

Fax: (772) 462-5961

Email: mbausher@ushrl.ars.usda.gov

Seq primer: T3 Primer.

## FEATURES

source

1..554 Location/Qualifiers

/organism="Citrus sinensis"

/mol\_type="mRNA"

/cultivar="Ridge Pineapple"

/db\_xref="taxon:2711"

/clone="USDA-FP\_00220"

/tissue\_type="entire seedling"

/dev\_stage="50 days after germination"

/lab\_host="XL1-Blue"

/clone\_lib="Ridge pineapple sweet orange entire seedling"

/note="Vector: pBluescript II SK+; Site 1: BcoRI; Site 2: XhoI; A high quality EST with at least 200 contiguous bases at Trace Tumor score of 20 or better"

## ORIGIN

Query Match 71.5%; Score 18.6; DB 13; Length 554;  
Best Local Similarity 84.0%; Pred. No. 3.4e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26

Db 147 CCACGGCGAGCGCCGAATCAACGAA 123

## RESULT 13

CA140644/c

LOCUS

DEFINITION SCJFRT2053G09.b RP2 Saccharum officinarum cDNA clone SCJFRT2053G09 3', mRNA sequence.

ACCESSION CA140644

VERSION CA140644.1 GI:35033841

KEYWORDS  
SOURCE  
ORGANISM

EST.  
Saccharum officinarum  
Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; FACCAD clade; Panicoideae; Andropogoneae; Saccharum.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 564) Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at

http://www.bcccenter.fcav.unesp.br

Plate: 053 row: G column: 09

Seq primer: SP6 Promoter primer.

## FEATURES

source

1..564 Location/Qualifiers

/organism="Saccharum officinarum"

/mol\_type="mRNA"

/db\_xref="taxon:4547"

/clone="SCJFRT2053G09"

/lab\_host="DH10B"

/clone\_lib="RT2"

/note="Organ: Root tips (0.3cm-long) from adult plants; Vector: pSPort1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Root tips (0.3cm-long) from adult plants]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at

http://sucest.lad.ic.unicamp.br/public"

## ORIGIN

Query Match 71.5%; Score 18.6; DB 13; Length 564;  
Best Local Similarity 84.0%; Pred. No. 3.4e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGAATCTACGAA 26

Db 482 CAACGGCGAGCGCCGTACATCTACGAA 458

## RESULT 14

CF507469/c

LOCUS

DEFINITION

CF507469

USDA-FP 122000-970 Immature Ovaries from field-collected Valencia Sweet Orange (Citrus sinensis (L.) Osbeck) Citrus sinensis cDNA

clone MFP-57\_H05 5', mRNA sequence.

CF507469

CF507469.1 GI:34522653

EST.

SOURCE Citrus sinensis

ORGANISM Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Sapindales; Rutaceae; Citrus.

1 (bases 1 to 610) Chaparro, J.X., Bausher, M.G., Dang, P., Hunter, W.B., McKenzie, C.L., Niedz, R.P. and Shatters, R.G. Jr.

Expressed Sequence Tags from Immature Ovaries of Valencia Sweet Orange (Citrus sinensis (L.) Osbeck)

UNPUBLISHED (2003)

CONTACT: Chaparro, JX



Horticulture and Breeding Research Unit  
 USDA, ARS, U. S. Horticultural Research Laboratory  
 2001, South Rock Road, Fort Pierce, FL 34945, USA  
 Tel: 772 462 5830  
 Fax: 772 462 5986  
 Email: jchaparro@ushrl.ars.usda.gov  
 Seq primer: T3 Primer

# FEATURES

source  
 1. .610  
 /organism="Citrus sinensis"  
 /mol\_type="mRNA"  
 /cultivar="Valencia"  
 /db\_xref="taxon:2711"  
 /clone="MVP-57 H05"  
 /tissue\_type="Immature ovaries"  
 /clone\_lib="Immature Ovaries from field-collected Valencia  
 Sweet Orange (Citrus sinensis (L.) Osbeck)"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; Standard library construction protocols from  
 StrataGene cDNA synthesis kit (cat.# 200401-5) and Uni-ZAP  
 XR vector kit (cat.# 237211) were followed using poly(A)  
 RNA."

# ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 610;  
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 CAACGGCGAGCCCGAATCTACGAA 26  
 Db 178 CCACGGCGAGCCCGAATAACGAA 154

# RESULT 15

CF653532/c  
 LOCUS  
 DEFINITION  
 USDA-FP\_003419 Ridge pineapple sweet orange entire seedling Citrus  
 sinensis cDNA clone RSE08E08 5', mRNA sequence.

ACCESSION  
 VERSION  
 CF653532.1 GI:37509341  
 EST.

SOURCE  
 Citrus sinensis

# ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Sapindales; Rutaceae; Citrus.  
 1 (bases 1 to 618)  
 Bausher, M., Shatters, R., Chaparro, J., Dang, P., Hunter, W. and  
 Niedz, R.

TITLE  
 An expressed sequence tag (EST) set from Citrus sinensis L. Osbeck  
 whole seedlings and the implications of further perennial source  
 investigations

# JOURNAL

COMMENT  
 Plant Sci. 165, 415-422 (2003)  
 Contact: Michael Bausher  
 US Horticultural Research

USDA - ARS  
 2001 South Rock Rd., Fort Pierce, FL 34945, USA  
 Tel: (772) 462-5918

Fax: (772) 462-5961  
 Email: mbausher@ushrl.ars.usda.gov

Seq primer: T3 Universal  
 Location/Qualifiers

# FEATURES

source  
 1. .618  
 /organism="Citrus sinensis"  
 /mol\_type="mRNA"  
 /cultivar="Ridge Pineapple"  
 /db\_xref="taxon:2711"  
 /clone="RSE08E08"  
 /tissue\_type="entire seedling"  
 /dev\_stage="50 Days after germination"  
 /lab\_host="Xf1-Blue"  
 /clone\_lib="Ridge pineapple sweet orange entire seedling"  
 /note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:

# ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 618;  
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 CAACGGCGAGCCCGAATCTACGAA 26  
 Db 200 CCACGGCGAGCCCGAATAACGAA 176

# RESULT 16

CF417287/c  
 LOCUS  
 DEFINITION  
 USDA-FP\_115000-217 Citrus sinensis; Insect-damaged immature  
 leaves/stems from field collected Valencia orange Citrus sinensis  
 cDNA clone IDFL-003\_E10 5', mRNA sequence.

ACCESSION  
 VERSION  
 CF417287.1 GI:34418414  
 EST.

SOURCE  
 Citrus sinensis

# ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Sapindales; Rutaceae; Citrus.  
 1 (bases 1 to 635)  
 Shatters, R.G. Jr., Chaparro, J.X., Bausher, M.G., Hunter, W.G. and  
 Niedz, R.P.

REFERENCE  
 AUTHORS  
 Expressed Sequence Tags from Insect-Damaged Field-Collected Sweet  
 Orange (Citrus Sinensis (L.) Osbeck) Immature Leaves and Stems

# TITLE

JOURNAL  
 COMMENT  
 Contact: Shatters, Jr. RG  
 Subtropical Insects Research Unit  
 USDA, ARS, U. S. Horticultural Research Laboratory  
 2001, South Rock Road, Fort Pierce, FL 34945, USA  
 Tel: 772 462 5912  
 Fax: 772 462 5986  
 Email: rshatters@ushrl.ars.usda.gov  
 Seq primer: T3 Primer;

# FEATURES

source  
 1. .635  
 /organism="Citrus sinensis"  
 /mol\_type="mRNA"  
 /cultivar="Valencia"  
 /isolate="mixed field population"  
 /db\_xref="taxon:2711"  
 /clone="IDFL-003\_E10"  
 /tissue\_type="Flush leaves and stems"  
 /clone\_lib="Citrus sinensis: Insect-damaged immature  
 leaves/stems from field collected Valencia orange"  
 /note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
 XhoI; Citrus sinensis (L.) Osbeck; Standard library  
 construction protocols from StrataGene cDNA synthesis kit  
 (cat.# 200401-5) and Uni-ZAP XR vector kit (cat.# 237211)  
 were followed using poly(A) RNA."

# ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 635;  
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 2 CAACGGCGAGCCCGAATCTACGAA 26  
 Db 165 CCACGGCGAGCCCGAATAACGAA 141

# RESULT 17

BH885920/c  
 LOCUS  
 DEFINITION  
 BH885920.d.sp6.1 Leishmania major Friedlin BAC Library Leishmania  
 major genomic clone LB00482a, genomic survey sequence.

```

ACCESSION      BH885920
VERSION        BH885920.1  GI:22130315
KEYWORDS       GSS.
SOURCE         Leishmania major
ORGANISM       Leishmania major
               Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
               Leishmania.
REFERENCE      1 (bases 1 to 653)
               Myler,P.J., Vogt,C., Munden,H., Robertson,L., Sisk,E.,
               Fazelinia,G., Aggarwal,G., Nelson,S., Seyler,A., Worthey,E.,
               Stuart,K. and Ragland,M.
               Leishmania major Friedlin BAC End Sequences
TITLE          Unpublished (2002)
JOURNAL        Other GSSs: LB00482a.d.T7.1
COMMENT        Contact: Myler PJ
               Seattle Biomedical Research Institute
               4 Nickerson Street, Seattle, WA 98109-1651, USA
               Tel: 206 284-8845
               Fax: 206 284-0313
               Email: mylerpj@sbri.org
               Seq primer: SP6
               Class: BAC ends.
FEATURES       Location/Qualifiers
               source
               1..653
               /organism="Leishmania major"
               /mol_type="genomic DNA"
               /strain="Friedlin"
               /db_xref="taxon:5664"
               /clone="LB00482a"
               /lab_host="E. coli GeneHogs + TrfA"
               /note="Vector: pCG270; Site 1: HindIII; Genomic DNA from
               Leishmania major Friedlin in agarose blocks was partially
               digested with HindIII, size selected, and ligated with
               HindIII-digested pCG270 vector DNA. 10368 clones were
               picked and arrayed in 384- and 96-well plates. Library
               construction and arraying was carried out by ResGen
               Corporation and clones and filters are available from
               them"
ORIGIN
Query Match      71.5%; Score 18.6; DB 28; Length 653;
Best Local Similarity 84.0%; Pred. No. 3.5e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGCGAGGCCGCAATCTACGAA 26
Db 372 CACGCGAGGCCGCAATCTACGAA 348

RESULT 18
LOCUS      CF417889/c
DEFINITION USDA-FP_115000-819 Citrus sinensis: Insect-damaged immature
leaves/stems from field collected Valencia orange Citrus sinensis
cDNA clone IDFL-011_C09 5', mRNA sequence.
ACCESSION  CF417889
VERSION     CF417889.1  GI:34419016
KEYWORDS    EST.
SOURCE      Citrus sinensis
ORGANISM    Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE   1 (bases 1 to 656)
AUTHORS     Shatters, R.G. Jr., Chaparro,J.X., Bausher,M.G., Hunter,W.G. and
Niedz,R.P.
TITLE       Expressed Sequence Tags from Insect-Damaged Field-Collected Sweet
Orange (Citrus Sinensis (L.) Osbeck) Immature Leaves and Stems
JOURNAL     Unpublished (2003)
COMMENT     Contact: Shatters, Jr. RG
             Subtropical Insects Research Unit
             USDA, ARS, U. S. Horticultural Research Laboratory

CF417889      656 bp  mRNA  linear  EST 02-SEP-2003
USDA-FP_115000-819 Citrus sinensis: Insect-damaged immature
leaves/stems from field collected Valencia orange Citrus sinensis
cDNA clone IDFL-011_C09 5', mRNA sequence.
CF417889
CF417889.1  GI:34419016
EST.
Citrus sinensis
Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE   1 (bases 1 to 656)
AUTHORS     Shatters, R.G. Jr., Chaparro,J.X., Bausher,M.G., Hunter,W.G. and
Niedz,R.P.
TITLE       Expressed Sequence Tags from Insect-Damaged Field-Collected Sweet
Orange (Citrus Sinensis (L.) Osbeck) Immature Leaves and Stems
JOURNAL     Unpublished (2003)
COMMENT     Contact: Shatters, Jr. RG
             Subtropical Insects Research Unit
             USDA, ARS, U. S. Horticultural Research Laboratory

CF417889      656 bp  mRNA  linear  EST 02-SEP-2003
USDA-FP_115000-819 Citrus sinensis: Insect-damaged immature
leaves/stems from field collected Valencia orange Citrus sinensis
cDNA clone IDFL-011_C09 5', mRNA sequence.
CF417889
CF417889.1  GI:34419016
EST.
Citrus sinensis
Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE   1 (bases 1 to 656)
AUTHORS     Shatters, R.G. Jr., Chaparro,J.X., Bausher,M.G., Hunter,W.G. and
Niedz,R.P.
TITLE       Expressed Sequence Tags from Insect-Damaged Field-Collected Sweet
Orange (Citrus Sinensis (L.) Osbeck) Immature Leaves and Stems
JOURNAL     Unpublished (2003)
COMMENT     Contact: Shatters, Jr. RG
             Subtropical Insects Research Unit
             USDA, ARS, U. S. Horticultural Research Laboratory

CF417889      656 bp  mRNA  linear  EST 01-JUL-2002
USDA-FP_00454 Ridge pineapple sweet orange entire seedling Citrus
sinensis cDNA clone USDA-FP_00454 5', mRNA sequence.
CF417889
CF417889.1  GI:21650532
EST.
Citrus sinensis
Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
REFERENCE   1 (bases 1 to 679)
AUTHORS     Bausher,M., McKendree,W., Dang,P., Chaparro,J., Shatters,R.,
Hunter,W. and Niedz,R.
TITLE       Expressed sequence tags isolated from entire sweet orange (C.
sinensis L. Osbeck) seedling
JOURNAL     Unpublished (2003)
COMMENT     Contact: Michael Bausher
             US Horticultural Research
             USDA - ARS
             2001 South Rock Rd., Fort Pierce, FL 34945, USA
             Tel: (772) 462-5918
             Fax: (772) 462-5961
             Email: mbausher@ushrl.ars.usda.gov
             Seq primer: T3 Primer.
             Location/Qualifiers
             source
             1..679
             /organism="Citrus sinensis"
             /mol_type="mRNA"
             /cultivar="Ridge Pineapple"
             /db_xref="taxon:2711"
             /clone="USDA-FP_00454"
             /tissue_type="entire seedling"
             /dev_stage="50 days after germination"
             /lab_host="X11-Blue"
             /clone_lib="Ridge pineapple sweet orange entire seedling"
             /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site_2:
             XhoI; A high quality EST with at least 200 contiguous
             bases at Trace Tuner score of 20 or better"

```

2001, South Rock Road, Fort Pierce, FL 34945, USA

Tel: 772 462 5912

Fax: 772 462 5986

Email: rshatters@ushrl.ars.usda.gov

Seq primer: T3 Primer.

Location/Qualifiers

source

1..656

/organism="Citrus sinensis"

/mol\_type="mRNA"

/cultivar="Valencia"

/isolate="mixed field population"

/db\_xref="taxon:2711"

/clone="IDFL-011\_C09"

/tissue\_type="flush leaves and stems"

/clone\_lib="Citrus sinensis: Insect-damaged immature

leaves/stems from field collected Valencia orange"

/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site\_2:

XhoI; Citrus sinensis (L.) Osbeck; Standard library

construction protocols from Stratagene cDNA synthesis kit

(cat.# 200401-5) and Uni-ZAP XR vector kit (cat.# 237211)

were followed using poly(A) RNA."

#### ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 656;

Best Local Similarity 84.0%; Pred. No. 3.5e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGCGAGGCCGCAATCTACGAA 26

Db 132 CCACGCGAGGCCGCAATCTACGAA 108

#### RESULT 19

BO623363/c

LOCUS

DEFINITION

USDA-FP\_00454 Ridge pineapple sweet orange entire seedling Citrus

sinensis cDNA clone USDA-FP\_00454 5', mRNA sequence.

BO623363

VERSION

BO623363.1 GI:21650532

KEYWORDS

EST.

SOURCE

ORGANISM

Citrus sinensis

Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Citrus.

REFERENCE

1 (bases 1 to 679)

AUTHORS

Bausher,M., McKendree,W., Dang,P., Chaparro,J., Shatters,R.,

Hunter,W. and Niedz,R.

TITLE

Expressed sequence tags isolated from entire sweet orange (C.

sinensis L. Osbeck) seedling

JOURNAL

Unpublished (2003)

COMMENT

Contact: Michael Bausher

US Horticultural Research

USDA - ARS

2001 South Rock Rd., Fort Pierce, FL 34945, USA

Tel: (772) 462-5918

Fax: (772) 462-5961

Email: mbausher@ushrl.ars.usda.gov

Seq primer: T3 Primer.

Location/Qualifiers

source

1..679

/organism="Citrus sinensis"

/mol\_type="mRNA"

/cultivar="Ridge Pineapple"

/db\_xref="taxon:2711"

/clone="USDA-FP\_00454"

/tissue\_type="entire seedling"

/dev\_stage="50 days after germination"

/lab\_host="X11-Blue"

/clone\_lib="Ridge pineapple sweet orange entire seedling"

/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site\_2:

XhoI; A high quality EST with at least 200 contiguous

bases at Trace Tuner score of 20 or better"

## ORIGIN

Query Match 71.5%; Score 18.6; DB 13; Length 679;  
 Best Local Similarity 84.0%; Pred. No. 3.5e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTACGAA 26  
 Db 198 CCACGCCGAGCCCGCAAAATACGAA 174

## RESULT 20

CF837440 UCRS03\_04K20\_r Washington Navel Orange Shoot Meristem cDNA Library EST 30-OCT-2003  
 LOCUS UCRS03\_04K20\_r Washington Navel Orange Shoot Meristem cDNA Library  
 DEFINITION Citrus sinensis cDNA clone CS\_PEA04K20, mRNA sequence.

## ACCESSION

CF837440

## VERSION

CF837440.1 GI:38053092

## KEYWORDS

EST.

## SOURCE

Citrus sinensis

Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Citrus.

1 (bases 1 to 691)

Close,T.J., Roose,M.L., Federici,C.F., Mu,L., Fenton,R.D.,

Wanamaker,S., Kim,H.R., Kudrna,D., Wing,R. and Yu,Y.

Development of EST Resources and New Genetic Markers for California

Citrus - Washington Navel Orange Shoot Meristem

Unpublished (2003)

Contact: Timothy Close

Department of Botany & Plant Sciences, University of California

Riverside, CA, 92521-0124

Tel: 9097873318

Fax: 9097874437

Email: timothy.close@ucr.edu

Seq primer: T3.

Location/Qualifiers

1..691

/organism="Citrus sinensis"

/mol\_type="mRNA"

/cultivar="Parent Washington Navel"

/db\_xref="taxon:2711"

/clone="CS\_PEA04K20"

/tissue\_type="Shoot meristem"

/dev\_stage="10 year old trees"

/lab\_host="E. coli TJC121"

/clone\_lib="Washington Navel Orange Shoot Meristem cDNA

Library"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site 1: ECoRI; Site 2: XhoI; Parent Washington Navel

Orange trees on Troyer rootstock (UCR 16K) were the source

of tissue. Trees at UC Riverside Agricultural Operations,

were planted October 12, 1992. In each of 17 reps one tree

on Troyer rootstock was initially treated with Enzone, one

with Alliette and Nemacure, and one was left untreated.

These treatments were discontinued in 1998. At the time of

sampling, there were differences in the apparent health

and size of the trees on Troyer rootstock. Fall-flush

shoots were sampled in early November 2002 to minimize the

number of floral shoot meristems. Federici and Mu (Roose

lab) harvested meristems only from trees that appeared to

be healthy and had a large number of young shoot tips on

the day of collection. The average weight of a meristem

was about 2 mg. Federici noted that there were quite a few

insects and signs of insect damage to the shoot tips.

Mealy bugs, thrips and aphids were observed, plus a few

very tiny fast moving insects that may have been mites or

crawler stage of scale (although Federici did not see any

mature scale). It was not difficult to avoid collecting

most of these because they were easy to see with the

dissecting microscope. It was harder to exclude the frass.

Some frass was definitely retained in the samples. Tissues

## ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 691;  
 Best Local Similarity 84.0%; Pred. No. 3.6e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGCCCGCAATCTACGAA 26

Db 209 CCACGCCGAGCCCGCAAAATACGAA 185

## RESULT 21

CF653361/c

LOCUS

DEFINITION

sinensis cDNA clone RSE05C05 5', mRNA sequence.

CF653361

CF653361.1 GI:37509170

EST.

Citrus sinensis

Citrus sinensis

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Sapindales; Rutaceae; Citrus.

1 (bases 1 to 692)

Bausher,M., Shatters,R., Chaparro,J., Dang,P., Hunter,W. and

Niedz,R.

An expressed sequence tag (EST) set from Citrus sinensis L. Osbeck

whole seedlings and the implications of further perennial source

investigations

Plant Sci. 165, 415-422 (2003)

Contact: Michael Bausher

US Horticultural Research

USDA - ARS

2001 South Rock Rd., Fort Pierce, FL 34945, USA

Tel: (772) 462-5918

Fax: (772) 462-5961

Email: mbausher@ushrl.ars.usda.gov

Seq primer: T3 Universal.

Location/Qualifiers

1..692

/organism="Citrus sinensis"

/mol\_type="mRNA"

/cultivar="Ridge Pineapple"

/db\_xref="taxon:2711"

/clone="RSE05C05"

/tissue\_type="entire seedling"

/dev\_stage="50 days after germination"

/lab\_host="XhI-Blue"

/clone\_lib="Ridge pineapple sweet orange entire seedling"

/note="Vector: pBluescript II SK+; Site 1: ECoRI; Site 2:

XhoI; A high quality EST with at least 200 contiguous

bases at Trace Tunes score of 20 or better"

## ORIGIN

were snap frozen and then stored at -80C until further processing. Fenton (Close lab) purified RNA by the phenol method described in J. Japanese Soc. Hort. Sci. 1996. 64 (4): 809-814, purified poly(A) mRNA using a PolyAtrack mRNA Isolation System IV (Promega), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised one million pfu from the primary library to produce a phagemid population. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Kim, Kudrna, Wing, Yu). Chromatogram files were downloaded by FTP to UC Riverside (by Close), then processed at UC Riverside (by Wanamaker, Close lab) using the HarVest pipeline (http://harvest.ucr.edu) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."



<p>Best Local Similarity 84.0%; Pred. No. 3.6e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;</p>		<p>Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."</p>	
QY	<p>2 CAACGCCGAGGCCGGAATCTACGAA 26</p> <p>     </p>	<p>Query Match 71.5%; Score 18.6; DB 14; Length 769; Best Local Similarity 84.0%; Pred. No. 3.6e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;</p>	<p>ORIGIN</p>
Db	<p>629 CCACGCCGAGGCCGGAATCTACGAA 653</p> <p>     </p>		
RESULT 24			
CD574354	769 bp mRNA linear EST 12-JUN-2003		
LOCUS	UCRPT01_05_E12_TV Poncirus trifoliata CTV-challenged cDNA library -		
DEFINITION	UCR Poncirus trifoliata cDNA clone UCRPT01_05_E12, mRNA sequence.		
ACCESSION	CD574354		
VERSION	CD574354.1 GI:31670256		
KEYWORDS	EST.		
SOURCE	Poncirus trifoliata		
ORGANISM	Poncirus trifoliata		
REFERENCE	<p>Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D., Wanamaker,S., Choi,Y. and Kingan,T. Development of EST Resources and New Genetic Markers for California Citrus - Poncirus trifoliata CTV-challenged phloem - UCR Unpublished (2003) Contact: Mikeal Roose Department of Botany &amp; Plant Sciences, University of California Riverside, CA, 92521-0124, USA Tel: 9097874137 Fax: 9097874437 Email: mikeal.roose@ucr.edu Seq primer: (T)20V.</p>		
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES	<p>source</p> <p>1..769</p> <p>Location/Qualifiers</p> <p>/organism="Poncirus trifoliata"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Pomeroy Op"</p> <p>/db_xref="taxon:37690"</p> <p>/clone="UCRPT01_05_E12"</p> <p>/tissue_type="Phloem"</p> <p>/dev_stage="10 - 30 cm shoots"</p> <p>/lab_host="E. coli TUC121"</p> <p>/clone_lib="Poncirus trifoliata CTV-challenged cDNA library - UCR"</p> <p>/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; plants were grown in the greenhouse at University of California, Riverside. The scion was a open-pollinated (very probably selfed) seedling of Poncirus trifoliata cv Pomeroy that was selected as homozygous for the Ctv resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate T514 over 1 year before sampling (CTV infests sweet orange, but not genotypes carrying the Ctv resistance gene. Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the University of California Riverside Genomics Institute, the Instrumentation Facility. (Choi, Kingan). Chromatogram files were transmitted to UC Riverside (by Choi), then processed at UC Riverside (by Wanamaker) using the HarVEST pipeline (<a href="http://harvest.ucr.edu">http://harvest.ucr.edu</a>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region.</p>		
QY	<p>2 CAACGCCGAGGCCGGAATCTACGAA 26</p> <p>     </p>		
Db	<p>629 CCACGCCGAGGCCGGAATCTACGAA 653</p> <p>     </p>		
RESULT 25			
CD574353	799 bp mRNA linear EST 12-JUN-2003		
LOCUS	UCRPT01_05_E12_T3 Poncirus trifoliata CTV-challenged cDNA library -		
DEFINITION	UCR Poncirus trifoliata cDNA clone UCRPT01_05_E12, mRNA sequence.		
ACCESSION	CD574353		
VERSION	CD574353.1 GI:31670255		
KEYWORDS	EST.		
SOURCE	Poncirus trifoliata		
ORGANISM	Poncirus trifoliata		
REFERENCE	<p>Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D., Wanamaker,S., Choi,Y. and Kingan,T. Development of EST Resources and New Genetic Markers for California Citrus - Poncirus trifoliata CTV-challenged phloem - UCR Unpublished (2003) Contact: Mikeal Roose Department of Botany &amp; Plant Sciences, University of California Riverside, CA, 92521-0124, USA Tel: 9097874137 Email: mikeal.roose@ucr.edu Seq primer: T3.</p>		
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES	<p>source</p> <p>1..799</p> <p>Location/Qualifiers</p> <p>/organism="Poncirus trifoliata"</p> <p>/mol_type="mRNA"</p> <p>/cultivar="Pomeroy Op"</p> <p>/db_xref="taxon:37690"</p> <p>/clone="UCRPT01_05_E12"</p> <p>/tissue_type="Phloem"</p> <p>/dev_stage="10 - 30 cm shoots"</p> <p>/lab_host="E. coli TUC121"</p> <p>/clone_lib="Poncirus trifoliata CTV-challenged cDNA library - UCR"</p> <p>/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; plants were grown in the greenhouse at University of California, Riverside. The scion was a open-pollinated (very probably selfed) seedling of Poncirus trifoliata cv Pomeroy that was selected as homozygous for the Ctv resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate T514 over 1 year before sampling (CTV infests sweet orange, but not genotypes carrying the Ctv resistance gene. Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the University of California Riverside Genomics Institute, the Instrumentation Facility. (Choi, Kingan). Chromatogram files were transmitted to UC Riverside (by Choi), then processed at UC Riverside (by Wanamaker) using the HarVEST pipeline (<a href="http://harvest.ucr.edu">http://harvest.ucr.edu</a>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region.</p>		

Chromatogram files were transmitted to UC Riverside (by Choi), then processed at UC Riverside (by Wananaker) using the HarVest pipeline (<http://harvest.ucr.edu>) using vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 71.5%; Score 18.6; DB 14; Length 799;  
Best Local Similarity 84.0%; Pred. No. 3.7e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGGAATCTACGAA 26  
DB 184 CCACGGCGAGGCCGGAATCTACGAA 160

RESULT 26  
BF513871/c  
LOCUS  
DEFINITION  
UI-H-BM1-amq-b-02-0-UI-s1 NCI CGAP\_Sub7 Homo sapiens cDNA clone  
IMAGE:3070778 3', mRNA sequence.  
ACCESSION  
BF513871  
VERSION  
BF513871.1 GI:11599050  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 468)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Oligo-dt track not found. Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: NCI-CGAP clone distribution  
Information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html) The following repetitive  
elements were found in this cDNA sequence: 18-43,  
>GC-rich#Low complexity  
Seq primer: M13 Forward  
POLYA=No.

## FEATURES

source

Location/Qualifiers  
1..468  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3070778"  
/lab\_host="NCI CGAP\_Sub7"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub7  
is a subcloned library derived from NCI CGAP Sub6. The  
NCI CGAP Sub7 library had 12 million recombinants. A  
single-stranded DNA preparation of NCI CGAP Sub6 was used  
as a tracer in a subtractive hybridization with a driver  
comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM  
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones  
1322376-1323911, 1456008-1456775, 1500552-1502855);  
NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778  
(IMAGE Clones 1323912-1325831, 1471368-1472903,  
1492104-1493255); NCI CGAP\_Lus pool 1 LLAM 3575-3582,  
3851-3854 (IMAGE Clones 1414920-1417991,  
1520904-1522439); NCI CGAP\_GC4 pool 1 LLAM 3164-3167,  
3716-3720, 3733-3735 (IMAGE Clones  
1257096-1258631, 1469064-1470983, 1475592-1476743);  
NCI CGAP\_Pt22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068  
(IMAGE Clones 985608-986759, 1101192-1101959,  
1217928-1220615); NCI CGAP\_Col0 pool 1 LLAM 2644-2653,

## ORIGIN

Query Match 70.0%; Score 18.2; DB 10; Length 468;  
Best Local Similarity 87.0%; Pred. No. 4.7e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCGGAATCTAC 23  
DB 460 ACAACGGCGAGGCCGGAATCTAC 438

## RESULT 27

BF515709/c  
LOCUS  
DEFINITION  
UI-H-BM1-amq-b-02-0-UI-s1 NCI CGAP\_Sub7 Homo sapiens cDNA clone  
IMAGE:3082707 3', mRNA sequence.  
ACCESSION  
BF515709  
VERSION  
BF515709.1 GI:11600888  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 486)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Oligo-dt track not found. Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: NCI-CGAP clone distribution  
Information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html) The following repetitive  
elements were found in this cDNA sequence: 18-43,  
>GC-rich#Low complexity  
Seq primer: M13 Forward  
POLYA=No.

## FEATURES

source

Location/Qualifiers  
1..486  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3082707"  
/lab\_host="NCI CGAP\_Sub7"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub7  
is a subcloned library derived from NCI CGAP Sub6. The  
NCI CGAP Sub7 library had 12 million recombinants. A  
single-stranded DNA preparation of NCI CGAP Sub6 was used  
as a tracer in a subtractive hybridization with a driver  
comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM  
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones  
1322376-1323911, 1456008-1456775, 1500552-1502855);  
NCI CGAP\_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778

(IMAGE Clonoids 1323912-1325831, 1471368-1472903, 1492104-1493255); NCI CGAP LAM5 pool 1 LHAM 3575-3582, 3851-3854 (IMAGE Clonoids 1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LHAM 3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids 1257096-1258631, 1469064-1470983, 1475592-1476743); NCI CGAP Pz22 pool 1 LHAM 2457-2459, 2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759, 1101192-1101959, 1217928-1220615); NCI CGAP\_Co10 pool 1 LHAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255, 1144584-1145351). (6% of the driver population), plus a pool of 3,840 arrayed clones from NCI CGAP\_Sub1 (IMAGE Clonoids 2708616-2710535) and NCI CGAP\_Sub2 (IMAGE Clonoids 2710536-2712455) (4% of the driver population), plus a pool of 11,136 clones from NCI CGAP\_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the driver population), plus a pool of 5,472 clones from NCI CGAP\_Sub4 (IMAGE Clonoids 2723592-2729326) (40% of the driver population), plus a pool of 4032 clones from NCI CGAP\_Sub6 (IMAGE Clonoids 2728969-2733190) (40% of the driver population). Subtraction was performed as previously described (Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806. TAG\_SEQ=None found"

## ORIGIN

Query Match 70.0%; Score 18.2; DB 10; Length 486;  
Best Local Similarity 87.0%; Pred. No. 4.7e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACAACGGGAGCGCCGAATCTAC 23

Db 460 AACACGGAGAGTCAGAACTCTAC 438

RESULT 28  
CE189676  
LOCUS  
DEFINITION  
tigr-gss-dog-17000371445035 Dog Library Canis familiaris genomic,  
genomic survey sequence.

VERSION  
CE189676

KEYWORDS  
GSS.

SOURCE  
Canis familiaris (dog)

ORGANISM  
Canis familiaris

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1 (Bases 1 to 538)  
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

Ruscher, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and  
Venter, J.C.

The dog genome: survey sequencing and comparative analysis

Science 301 (5641), 1898-1903 (2003)

TITLE  
22875432

MEDLINE  
14512627

COMMENT  
Contact: Kirkness EF

The Institute for Genomic Research

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,

Rockville, MD 20850, USA

Tel: 301-838-0200

Fax: 301-838-0208

Email: ekirknes@tigr.org

Class: shotgun.

## FEATURES

source

Location/Qualifiers

1..538

/organism="Canis familiaris"

/mol\_type="genomic DNA"

/strain="Standard Poodle"

/db\_xref="taxon:9615"

/clone\_lib="Dog Library"

/note="Site 1: BstXI; Libraries were prepared from

peripheral blood"

## ORIGIN

Query Match 70.0%; Score 18.2; DB 29; Length 538;  
Best Local Similarity 87.0%; Pred. No. 4.9e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ACAACGGGAGCGCCGAATCTAC 23

Db 466 ACAACGGGAGCGCCGAATCTAC 488

## RESULT 29

CA182954/c

LOCUS

DEFINITION

5', mRNA sequence.

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 703)

Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

http://www.bccc.net.br

Plate: 151 row: A column: 02

Seq primer: T7 Promoter Primer.

Location/Qualifiers

1..703

/organism="Saccharum officinarum"

/mol\_type="mRNA"

/db\_xref="taxon:4547"

/clone="SCEZST3151A02"

/lab\_host="DH10B"

/clone\_lib="ST3"

/note="Organ: Fourth apical stalk internodes of adult

plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An

unidirectional cDNA library generated from Fourth apical

stalk internodes of adult plants]. cDNA was prepared from

polyA+ mRNA using SuperScript Plasmid System Kit

(Invitrogen). The double-strand cDNAs were fractionated

in a sepharose CL-2B 40cm-columns and fragments sizing

between 0.8 and 1.5 kb were directionally cloned into the

vector. Details of each source of RNA and library

construction can be obtained at

http://sucet.lad.ic.unicamp.br/public"

## ORIGIN

Query Match 70.0%; Score 18.2; DB 14; Length 703;  
Best Local Similarity 87.0%; Pred. No. 5.2e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AACGGGAGCGCCGAATCTACGA 25

Db 272 AACGACGACGACCGAATCTACGA 250

## RESULT 30

BE469759

LOCUS

BE469759 735 bp mRNA linear EST 28-JUL-2000

DEFINITION IpHk02105 Head kidney cDNA library Ictalurus punctatus cDNA 5',  
mRNA sequence.  
ACCESSION BE469759  
VERSION BE469759.1 GI:9560250  
KEYWORDS EST  
SOURCE Ictalurus punctatus (channel catfish)  
ORGANISM Ictalurus punctatus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;  
Ictaluridae; Ictalurus.  
REFERENCE 1 (bases 1 to 735)  
AUTHORS Cao, D., Kocabas, A., Ju, Z., Karsi, A., Li, P., Patterson, A. and  
Liu, Z.J.  
TITLE Transcriptome analysis of channel catfish (Ictalurus punctatus):  
Gene cataloging and profiling from the head kidney  
JOURNAL Unpublished (2000)  
COMMENT The Fish Molecular Genetics and Biotechnology Laboratory,  
Department of Fisheries and Allied Aquacultures and Program of Cell  
and Molecular Biosciences  
Auburn University  
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA  
Tel: 334 844 4054  
Fax: 334 844 9208  
Email: zliu@acesag.auburn.edu  
Seq primer: M13 Reverse  
FEATURES  
source  
1..735  
/organism="Ictalurus punctatus"  
/mol\_type="mRNA"  
/db\_xref="taxon:7998"  
/clone\_lib="Head kidney cDNA library"  
/note="Organ: Head kidney; Vector: pSport1; Site\_1: NotI;  
Site\_2: SalI"  
ORIGIN  
Query Match 70.0%; Score 18.2; DB 10; Length 735;  
Best Local Similarity 87.0%; Pred. No. 5.2e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 4 ACGGGAGGCCCGATCTACGAA 26  
|||||  
Db 591 ACGGGGTGGTCCGTATCTACGAA 613  
Search completed: June 20, 2004, 14:14:45  
Job time : 1240.98 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 28.7513 Seconds  
(without alignments)  
501.846 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26  
Sequence: 1 acaacgcgagccgaatctacgaa 26

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

Issued Patents NA:\*\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq\*

2: /cgn2\_6/ptodata/2/ina/5B COMB.seq\*

3: /cgn2\_6/ptodata/2/ina/6A COMB.seq\*

4: /cgn2\_6/ptodata/2/ina/6B COMB.seq\*

5: /cgn2\_6/ptodata/2/ina/PCRU COMB.seq\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	26	100.0	401	3	US-08-990-823-75
C 2	26	100.0	401	4	US-09-477-135A-75
C 3	26	100.0	4403765	3	US-09-103-840A-2
C 4	26	100.0	4411529	3	US-09-103-840A-1
C 5	18.6	71.5	505	4	US-09-221-017B-376
C 6	17.6	67.7	536165	4	US-09-214-808-1
C 7	17	65.4	725	3	US-08-998-416-808
C 8	17	65.4	879	4	US-09-408-020-37
C 9	17	65.4	1200	4	US-09-252-991A-3224
C 10	17	65.4	1203	4	US-09-252-991A-3421
C 11	17	65.4	1272	4	US-09-252-991A-15250
C 12	17	65.4	1614	4	US-09-252-991A-15438
C 13	17	65.4	1644	4	US-09-252-991A-792
C 14	17	65.4	1845	4	US-09-252-991A-3262
C 15	17	65.4	2907	4	US-09-252-991A-831
C 16	17	65.4	42432	4	US-09-408-020-2
C 17	16.6	63.8	762	4	US-09-489-039A-2899
C 18	16.6	63.8	1305	4	US-09-489-039A-5991
C 19	16.6	63.8	1350	4	US-09-252-991A-1564
C 20	16.6	63.8	1494	4	US-09-252-991A-1456
C 21	16.6	63.8	1536	4	US-09-252-991A-1567
C 22	16.4	63.1	438	4	US-08-252-991A-15918
C 23	16.4	63.1	546	4	US-09-429-906B-1
C 24	16.4	63.1	762	4	US-09-252-991A-15888
C 25	16.4	63.1	885	4	US-09-252-991A-15774
C 26	16.4	63.1	1068	4	US-09-252-991A-15804
C 27	16.2	62.3	3063	4	US-09-252-991A-12299

28	16.2	62.3	3105	4	US-09-252-991A-12350
C 29	16.2	62.3	4403765	3	US-09-103-840A-2
C 30	16.2	62.3	4411529	3	US-09-103-840A-1
C 31	16	61.5	420	1	US-08-470-179-148
C 32	16	61.5	489	4	US-09-894-844-63
C 33	16	61.5	696	4	US-09-252-991A-12599
C 34	16	61.5	971	4	US-08-630-915A-197
C 35	16	61.5	1457	3	US-09-444-053-3
C 36	16	61.5	4079	4	US-09-016-434-1248
C 37	16	61.5	4519	4	US-09-023-685-1202
C 38	16	61.5	15378	3	US-08-785-420-1
C 39	15.6	60.0	423	1	US-08-470-179-107
C 40	15.6	60.0	423	1	US-08-470-179-133
C 41	15.6	60.0	423	1	US-08-470-179-151
C 42	15.6	60.0	426	1	US-08-470-179-179
C 43	15.6	60.0	456	4	US-09-325-932A-14
C 44	15.6	60.0	861	4	US-09-252-991A-15703
C 45	15.6	60.0	891	4	US-09-252-991A-15676
C 46	15.6	60.0	11050	4	US-10-204-708-85
C 47	15.4	59.2	294	1	US-08-221-817-9
C 48	15.4	59.2	294	1	US-08-454-439-9
C 49	15.4	59.2	294	5	PCT-US94-10487-9
C 50	15.4	59.2	294	5	PCT-US94-10487-9
C 51	15.4	59.2	346	3	US-08-998-416-68
C 52	15.4	59.2	346	3	US-08-998-416-71
C 53	15.4	59.2	360	4	US-09-252-991A-15436
C 54	15.4	59.2	474	4	US-09-252-991A-6449
C 55	15.4	59.2	531	4	US-09-252-991A-8239
C 56	15.4	59.2	642	4	US-09-252-991A-12174
C 57	15.4	59.2	879	4	US-09-408-020-71
C 58	15.4	59.2	1059	4	US-09-489-039A-3376
C 59	15.4	59.2	1059	4	US-09-791-165-5
C 60	15.4	59.2	1059	4	US-09-791-165-6
C 61	15.4	59.2	1104	4	US-09-489-039A-198
C 62	15.4	59.2	1239	4	US-09-252-991A-8289
C 63	15.4	59.2	1314	4	US-09-495-050A-174
C 64	15.4	59.2	1335	4	US-09-252-991A-12238
C 65	15.4	59.2	1360	4	US-09-602-472A-14
C 66	15.4	59.2	1671	4	US-09-252-991A-6099
C 67	15.4	59.2	1674	4	US-09-252-991A-6179
C 68	15.4	59.2	1688	4	US-09-791-165-3
C 69	15.4	59.2	1700	4	US-09-514-521-2
C 70	15.4	59.2	1700	4	US-09-791-165-1
C 71	15.4	59.2	1743	4	US-09-252-991A-6372
C 72	15.4	59.2	1887	4	US-09-149-727-9
C 73	15.4	59.2	1931	4	US-09-429-906B-5
C 74	15.4	59.2	2364	4	US-09-107-532A-2366
C 75	15.4	59.2	2639	1	US-09-252-991A-1430
C 76	15.4	59.2	2639	6	US-07-952-817-8
C 77	15.4	59.2	3456	4	US-09-602-472A-12
C 78	15.4	59.2	3556	1	US-07-971-624E-1
C 79	15.4	59.2	3572	1	US-07-971-624E-2
C 80	15.4	59.2	3660	4	US-09-107-532A-2366
C 81	15.4	59.2	3843	4	US-09-252-991A-1430
C 82	15.4	59.2	3931	1	US-08-125-468-1
C 83	15.4	59.2	3931	2	US-08-474-933-1
C 84	15.4	59.2	3931	2	US-08-408-020-1
C 85	15.4	59.2	3931	2	US-08-408-020-1
C 86	15.4	59.2	3931	2	US-08-752-760A-1
C 87	15.2	58.5	346	3	US-09-060-756-317
C 88	15.2	58.5	346	4	US-09-670-314-317
C 89	15.2	58.5	676	4	US-09-470-191-49
C 90	15.2	58.5	942	4	US-09-252-991A-10909
C 91	15.2	58.5	945	4	US-09-252-991A-10966
C 92	15.2	58.5	1278	4	US-09-252-991A-11277
C 93	15.2	58.5	1806	4	US-09-252-991A-11213
C 94	15.2	58.5	2312	1	US-07-736-178C-1
C 95	15.2	58.5	2325	2	US-08-714-677-3
C 96	15.2	58.5	2325	2	US-08-393-540-3
C 97	15.2	58.5	2325	2	US-08-714-537-3
C 98	15.2	58.5	3073	2	US-08-714-677-11
C 99	15.2	58.5	3073	2	US-08-393-540-11
C 100	15.2	58.5	3073	2	US-08-714-537-11

Sequence 12350, A  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 148, Appl  
Sequence 63, Appl  
Sequence 12599, A  
Sequence 197, Appl  
Sequence 1248, Ap  
Sequence 1202, Ap  
Sequence 1, Appli  
Sequence 107, Appl  
Sequence 133, Appl  
Sequence 151, Appl  
Sequence 179, Appl  
Sequence 14, Appl  
Sequence 15703, A  
Sequence 15676, A  
Sequence 85, Appl  
Sequence 9, Appli  
Sequence 9, Appli  
Sequence 217, Appl  
Sequence 9, Appli  
Sequence 68, Appl  
Sequence 71, Appl  
Sequence 15436, A  
Sequence 6449, Ap  
Sequence 8239, Ap  
Sequence 12174, A  
Sequence 71, Appl  
Sequence 3376, Ap  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 198, Appl  
Sequence 8289, Ap  
Sequence 174, Ap  
Sequence 12338, A  
Sequence 12338, A  
Sequence 14, Appl  
Sequence 6099, Ap  
Sequence 6179, Ap  
Sequence 3, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 6372, Ap  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 12368, A  
Patent No. 5210025  
Sequence 12, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 2366, Ap  
Sequence 1430, Ap  
Sequence 1588, Ap  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 317, Appl  
Sequence 317, Appl  
Sequence 49, Appl  
Sequence 10309, A  
Sequence 10366, A  
Sequence 11277, A  
Sequence 11213, A  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 11, Appl  
Sequence 11, Appl  
Sequence 11, Appl



c 247 14.6 56.2 1458 4 US-09-252-991A-14586  
c 248 14.6 56.2 1551 4 US-09-489-039A-4788  
c 249 14.6 56.2 1674 4 US-09-252-991A-11512  
c 250 14.6 56.2 1845 4 US-09-252-991A-8187  
c 251 14.6 56.2 2074 4 US-09-690-942-7  
c 252 14.6 56.2 2167 3 US-08-743-637B-16  
c 253 14.6 56.2 2167 3 US-08-526-840B-16  
c 254 14.6 56.2 2310 3 US-09-036-987A-25  
c 255 14.6 56.2 2310 3 US-09-370-700-25  
c 256 14.6 56.2 2310 3 US-09-603-207-25  
c 257 14.6 56.2 2367 4 US-09-489-039A-5660  
c 258 14.6 56.2 2457 4 US-09-252-991A-15181  
c 259 14.6 56.2 3098 4 US-09-252-991A-7487  
c 260 14.6 56.2 3938 4 US-10-164-595-45  
c 261 14.6 56.2 4131 4 US-09-252-991A-12944  
c 262 14.6 56.2 4326 4 US-09-252-991A-11572  
c 263 14.6 56.2 4378 4 US-10-164-595-47  
c 264 14.6 56.2 4443 4 US-09-252-991A-11871  
c 265 14.6 56.2 4648 4 US-09-620-312D-464  
c 266 14.6 56.2 6131 1 US-07-732-242C-8  
c 267 14.6 56.2 6051 2 US-08-576-626A-2  
c 268 14.6 56.2 9775 4 US-09-647-540A-1  
c 269 14.6 56.2 9775 4 US-09-647-540A-3  
c 270 14.6 56.2 9775 4 US-09-647-540A-7  
c 271 14.6 56.2 9775 4 US-10-119-600-1  
c 272 14.6 56.2 9775 4 US-10-119-600-3  
c 273 14.6 56.2 9775 4 US-10-119-600-5  
c 274 14.6 56.2 9775 4 US-10-119-600-7  
c 275 14.6 56.2 9775 4 US-10-119-600-7  
c 276 14.6 56.2 9775 4 US-10-119-651-1  
c 277 14.6 56.2 9775 4 US-10-119-651-3  
c 278 14.6 56.2 9775 4 US-10-119-651-5  
c 279 14.6 56.2 9775 4 US-10-119-651-7  
c 280 14.6 56.2 34094 4 US-09-292-034-1  
c 281 14.6 56.2 36412 4 US-08-311-731A-132  
c 282 14.4 55.4 28 2 US-08-859-998-732  
c 283 14.4 55.4 28 4 US-09-225-328-732  
c 284 14.4 55.4 28 4 US-09-225-328-732  
c 285 14.4 55.4 110 4 US-09-621-976-9864  
c 286 14.4 55.4 186 4 US-09-313-294A-5677  
c 287 14.4 55.4 270 4 US-09-489-039A-1546  
c 288 14.4 55.4 273 4 US-09-252-991A-4560  
c 289 14.4 55.4 348 3 US-09-060-756-142  
c 290 14.4 55.4 348 4 US-09-670-314-142  
c 291 14.4 55.4 363 4 US-09-489-039A-3412  
c 292 14.4 55.4 371 3 US-09-060-756-199  
c 293 14.4 55.4 371 4 US-09-670-314-199  
c 294 14.4 55.4 396 1 US-08-470-179-181  
c 295 14.4 55.4 420 1 US-08-470-179-162  
c 296 14.4 55.4 423 1 US-08-470-179-175  
c 297 14.4 55.4 423 1 US-08-470-179-186  
c 298 14.4 55.4 423 1 US-08-470-179-189  
c 299 14.4 55.4 443 3 US-09-060-756-172  
c 300 14.4 55.4 443 4 US-09-670-314-172

## ALIGNMENTS

RESULT 1  
US-08-990-823-75/c  
; Sequence 75, Application US/08990823D  
; Patent No. 6228371  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 49086  
; CURRENT APPLICATION NUMBER: US/08/990,823D  
; CURRENT FILING DATE: 1997-12-15  
; EARLIER APPLICATION NUMBER: US 96/10375  
; EARLIER FILING DATE: 1996-06-14  
; EARLIER APPLICATION NUMBER: 60/000,254

; EARLIER FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 113  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 75  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: Modified base  
; OTHER INFORMATION: n represents a or g or c or t/u  
US-08-990-823-75

Query Match 100.0%; Score 26; DB 3; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.0024;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26  
Db 148 ACAACGGCGAGCCCGGAATCTACGAA 123

## RESULT 2

US-09-477-135A-75/c  
; Sequence 75, Application US/09477135A  
; Patent No. 6572865  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 52888  
; CURRENT APPLICATION NUMBER: US/09/477,135A  
; CURRENT FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08990823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 75  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: variation  
; LOCATION: (1)..(401)  
; OTHER INFORMATION: n = A, C, G, or T  
US-09-477-135A-75

Query Match 100.0%; Score 26; DB 4; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.0024;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26  
Db 148 ACAACGGCGAGCCCGGAATCTACGAA 123

## RESULT 3

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6234328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 100.0%; Score 26; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26  
DB 4219170 ACAACGGCGAGCCCGGAATCTACGAA 4219195

RESULT 4  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 100.0%; Score 26; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26  
DB 4226924 ACAACGGCGAGCCCGGAATCTACGAA 4226949

RESULT 5  
US-09-221-017B-376  
Sequence 376, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: ROSS, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b

Query Match 100.0%; Score 26; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 0.0052;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGGAATCTACGAA 26  
DB 4226924 ACAACGGCGAGCCCGGAATCTACGAA 4226949

RESULT 6  
US-09-214-808-1/c  
Sequence 1, Application US/09214808A  
Patent No. 6475793  
GENERAL INFORMATION:  
APPLICANT: Rosenthal, Andre  
APPLICANT: Freiberg, Christoph  
APPLICANT: Perret, Xavier Philippe  
APPLICANT: Broughton, William John  
TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic  
Patent No. 6475793  
TITLE OF INVENTION: Plasmid  
FILE REFERENCE: CARP0068  
CURRENT APPLICATION NUMBER: US/09/214,808A  
CURRENT FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: PCT/IB97/00950  
PRIOR FILING DATE: 1997-07-10  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 1  
LENGTH: 536165  
TYPE: DNA  
ORGANISM: Rhizobium  
US-09-214-808-1

Query Match 71.5%; Score 18.6; DB 4; Length 505;  
Best Local Similarity 84.0%; Pred. No. 7.3;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGGAATCTACGAA 26  
DB 336 CAACGGCGATGCGCGGAATCAAGAA 360

RESULT 6  
US-09-214-808-1/c  
Sequence 1, Application US/09214808A  
Patent No. 6475793  
GENERAL INFORMATION:  
APPLICANT: Rosenthal, Andre  
APPLICANT: Freiberg, Christoph  
APPLICANT: Perret, Xavier Philippe  
APPLICANT: Broughton, William John  
TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic  
Patent No. 6475793  
TITLE OF INVENTION: Plasmid  
FILE REFERENCE: CARP0068  
CURRENT APPLICATION NUMBER: US/09/214,808A  
CURRENT FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: PCT/IB97/00950  
PRIOR FILING DATE: 1997-07-10  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 1  
LENGTH: 536165  
TYPE: DNA  
ORGANISM: Rhizobium  
US-09-214-808-1

Query Match 71.5%; Score 18.6; DB 4; Length 505;  
Best Local Similarity 84.0%; Pred. No. 7.3;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGGAATCTACGAA 26  
DB 336 CAACGGCGATGCGCGGAATCAAGAA 360

Query Match 67.7%; Score 17.6; DB 4; Length 536165;  
Best Local Similarity 83.3%; Pred. No. 39;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGCCGGAATCTACGAA 26  
Db 85320 AAGGGCGGCGCCGAAACAACGAA 85297

## RESULT 7

US-08-998-416-808  
; Sequence 808, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Knechtie, Jurgin  
; APPLICANT: Knechtie, Philipp  
; APPLICANT: Reibischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPPII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/998,416  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Weigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGCI976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 808:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 725 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1515UP  
; US-08-998-416-808  
Query Match 65.4%; Score 17; DB 3; Length 725;  
Best Local Similarity 80.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AACACGGCGAGCCGGAATCTACGAA 25  
Db 472 AACACGGCGAGCGATGATGATGACGA 496

## RESULT 8

US-09-408-020-37/c

; Sequence 37, Application US/09408020  
; Patent No. 6632937  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schlieper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCORP.002A  
; CURRENT APPLICATION NUMBER: US/09/408,020  
; CURRENT FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 60/102,294  
; PRIOR FILING DATE: 1998-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37  
; LENGTH: 879  
; TYPE: DNA  
; ORGANISM: Cenarchaeum symbiosum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(879)  
US-09-408-020-37

Query Match 65.4%; Score 17; DB 4; Length 879;  
Best Local Similarity 80.0%; Pred. No. 44;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCGGAATCTACGAA 25  
Db 434 ACCTCGGAGAGCCCGAGTCTACAA 410

## RESULT 9

US-09-252-991A-3224  
; Sequence 3224, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3224  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3224

Query Match 65.4%; Score 17; DB 4; Length 1200;  
Best Local Similarity 80.0%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGGAATCTACGAA 26  
Db 925 CAGCGCGAGCGCGGCTTCTACGAA 949

## RESULT 10

US-09-252-991A-3421/c  
; Sequence 3421, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

Query Match 65.4%; Score 17; DB 4; Length 1203;  
Best Local Similarity 80.0%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;  
Query 2 CAACGGCGAGCCCGCAATCTACGAA 26  
Db 239 CAGCGCGAGCGCGGCTTCTACGAA 235  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 3421  
LENGTH: 1203  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3421

Query Match 65.4%; Score 17; DB 4; Length 1203;  
Best Local Similarity 80.0%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Query 2 CAACGGCGAGCCCGCAATCTACGAA 26  
Db 239 CAGCGCGAGCGCGGCTTCTACGAA 235

RESULT 11  
US-09-252-991A-15250/c  
Sequence 15250, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15250  
LENGTH: 1272  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15250

Query Match 65.4%; Score 17; DB 4; Length 1272;  
Best Local Similarity 80.0%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Query 2 CAACGGCGAGCCCGCAATCTACGAA 26  
Db 319 CAACGCCCTGGCCCGATATACGAA 295

RESULT 12  
US-09-252-991A-15438  
Sequence 15438, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 15438  
LENGTH: 1614  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15438

Query Match 65.4%; Score 17; DB 4; Length 1614;  
Best Local Similarity 80.0%; Pred. No. 46;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Query 2 CAACGGCGAGCCCGCAATCTACGAA 26  
Db 462 CAACGCCCTGGCCCGATATACGAA 486

RESULT 13  
US-09-252-991A-792/c  
Sequence 792, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 792  
LENGTH: 1644  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-792

Query Match 65.4%; Score 17; DB 4; Length 1644;  
Best Local Similarity 80.0%; Pred. No. 46;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Query 1 ACAACGGCGAGCCCGCAATCTACGA 25  
Db 869 ACAACGGCGAGCCCGGTCGACAA 845

RESULT 14  
US-09-252-991A-3262  
Sequence 3262, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 3262  
LENGTH: 1845  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3262

Query Match 65.4%; Score 17; DB 4; Length 1845;  
Best Local Similarity 80.0%; Pred. No. 46;  
Matches 20; Conservative 0; Mismatches 5; Indels 5; Gaps 0;

Query 2 CAACGGCGAGCCCGCAATCTACGAA 26  
Db 1695 CAGCGCGAGCGCGGCTTCTACGAA 1719

RESULT 15

US-09-252-991A-831  
; Sequence 831, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 831  
; LENGTH: 2907  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-831

Query Match 65.4%; Score 17; DB 4; Length 2907;  
Best Local Similarity 80.0%; Pred. No. 48;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCGCAATCTACGA 25  
DB 812 ACAACGGCGAGGCCGCAATCTACGA 836

RESULT 16  
US-09-408-020-2/c  
; Sequence 2, Application US/09408020  
; Patent No. 6632937  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCOEP.002A  
; CURRENT APPLICATION NUMBER: US/09/408,020  
; CURRENT FILING DATE: 1999-09-29  
; PRIOR FILING DATE: 1998-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 42432  
; TYPE: DNA  
; ORGANISM: Cenarchaeum symbiosum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)...(10421)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (10625)...(11434)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (11478)...(13046)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (13046)...(14620)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (23558)...(24862)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (24913)...(25728)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (26504)...(26881)  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: (29655)...(30491)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (34559)...(36067)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (37002)...(37403)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (37404)...(38282)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (39454)...(40572)  
US-09-408-020-2

Query Match 65.4%; Score 17; DB 4; Length 42432;  
Best Local Similarity 80.0%; Pred. No. 61;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCGCAATCTACGA 25  
DB 37837 ACCTCGGAGAGGCCGCAATCTACAA 37813

RESULT 17  
US-09-489-039A-2899  
; Sequence 2899, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 2899  
; LENGTH: 762  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-2899

Query Match 63.8%; Score 16.6; DB 4; Length 762;  
Best Local Similarity 82.6%; Pred. No. 66;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCGCAATCTACG 24  
DB 267 CAGCGCGAGGCCGCAATCTACG 289

RESULT 18  
US-09-489-039A-5991/c  
; Sequence 5991, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 5991  
; LENGTH: 1305  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-5991

```
Query Match          63.8%; Score 16.6; DB 4; Length 1305;
Best Local Similarity 82.6%; Pred. No. 69;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGGAATCTAC 23
   ||||| ||||| ||||| |||||
Db 1141 ACATCGGCGCGCCGGAATGTAC 1119

RESULT 19
US-09-252-991A-1664
; Sequence 1664, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1664
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1664

Query Match          63.8%; Score 16.6; DB 4; Length 1350;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGGAATCTACG 24
   ||||| ||||| ||||| |||||
Db 130 CAACGGCGAGCGGAGGATCTATG 152

RESULT 20
US-09-252-991A-1456/c
; Sequence 1456, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1456
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1456

Query Match          63.8%; Score 16.6; DB 4; Length 1494;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGGAATCTACG 24
   ||||| ||||| ||||| |||||
Db 682 CAACGGCGAGCGGAGGATCTATG 660

RESULT 21
US-09-252-991A-1567
```

```
; Sequence 1567, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1567
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1567

Query Match          63.8%; Score 16.6; DB 4; Length 1536;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGGAATCTACG 24
   ||||| ||||| ||||| |||||
Db 855 CAACGGCGAGCGGAGGATCTATG 877

RESULT 22
US-09-252-991A-15918/c
; Sequence 15918, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15918
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15918

Query Match          63.1%; Score 16.4; DB 4; Length 438;
Best Local Similarity 76.9%; Pred. No. 79;
Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCCGGAATCTACGAA 26
   ||||| ||||| ||||| |||||
Db 155 ACACGGCGAGCGCCGCGACCCAGGAA 130

RESULT 23
US-09-429-906B-1
; Sequence 1, Application US/09429906B
; Patent No. 6553099
; GENERAL INFORMATION:
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Orozco, Jr., Emil M.
; TITLE OF INVENTION: Plant UDP-Glucose Dehydrogenase
; FILE REFERENCE: BB1271 US NA
; CURRENT APPLICATION NUMBER: US/09/429.906B
; CURRENT FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/107,274
; PRIOR FILING DATE: 1998-11-05
```





```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12299
; LENGTH: 3063
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12299

Query Match      62.3%; Score 16.2; DB 4; Length 3063;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGCGAGGCCCGCAATCTACGA 25
Db 1264 CGCGAGGCCCGCACCTACGA 1284

RESULT 28
US-09-252-991A-12350
; Sequence 1, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12350
; LENGTH: 3105
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12350

Query Match      62.3%; Score 16.2; DB 4; Length 3105;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 CGCGAGGCCCGCAATCTACGA 25
Db 2007 CGCGAGGCCCGCACCTACGA 2027

RESULT 29
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-103-840A-2

Query Match      62.3%; Score 16.2; DB 3; Length 4403765;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGCAATCTA 22
Db 3575487 CAACGGCGTCCGCCCGCAATCGA 3575467

Search completed: June 20, 2004, 11:46:40
Job time : 42.8763 secs
```

```
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      62.3%; Score 16.2; DB 3; Length 4403765;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGCAATCTA 22
Db 3570341 CAACGGCGTCCGCCCGCAATCGA 3570321

RESULT 30
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      62.3%; Score 16.2; DB 3; Length 4411529;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGCAATCTA 22
Db 3575487 CAACGGCGTCCGCCCGCAATCGA 3575467

Search completed: June 20, 2004, 11:46:40
Job time : 42.8763 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 147.058 Seconds  
(without alignments)  
809.940 Million cell updates/sec

Title: US-10-624-714-15

Perfect score: 26  
Sequence: 1 acaacgcgagggccgaattacgaa 26

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/US06\_PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US05\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
C	1	26	100.0	401	9	US-09-956-634-75
C	2	26	100.0	401	10	US-09-997-182-75
C	3	26	100.0	401	10	US-09-997-181-75
	4	18.6	71.5	505	13	US-10-194-163-376
	5	18	69.2	6357	13	US-10-424-599-132268
	6	17.8	68.5	806	13	US-10-424-599-67461
	7	17.6	67.7	531	15	US-10-029-386-6930
	8	17.6	67.7	612	13	US-10-425-114-22881
	9	17.6	67.7	1017	13	US-10-282-122A-12032
	10	17.6	67.7	2205	16	US-10-369-493-44269
C	11	17.6	67.7	536165	10	US-09-939-964-1
	12	17	65.4	383	13	US-10-424-599-110822
	13	17	65.4	669	15	US-10-156-761-2378
	14	17	65.4	672	16	US-10-369-493-27924
						Sequence 75, Appl
						Sequence 75, Appl
						Sequence 376, Appl
						Sequence 132268,
						Sequence 67461, A
						Sequence 6930, Ap
						Sequence 22881, A
						Sequence 12032, A
						Sequence 44269, A
						Sequence 1, Appl
						Sequence 110822,
						Sequence 2378, Ap
						Sequence 27924, A

65.4	17	821	13	US-10-027-632-168321	Sequence 168321,
65.4	17	821	13	US-10-027-632-168322	Sequence 168322,
65.4	17	821	16	US-10-027-632-168321	Sequence 168321,
65.4	17	821	16	US-10-027-632-168322	Sequence 168322,
65.4	17	879	14	US-10-027-806-37	Sequence 37, Appl
65.4	17	879	14	US-10-034-823-37	Sequence 37, Appl
65.4	17	879	15	US-10-027-801-37	Sequence 37, Appl
65.4	17	879	15	US-10-029-120-37	Sequence 37, Appl
65.4	17	1173	13	US-10-282-122A-13773	Sequence 13773, A
65.4	17	1391	13	US-10-425-114-23453	Sequence 23453, A
65.4	17	1557	16	US-10-369-493-28653	Sequence 28653, A
65.4	17	1563	16	US-10-369-493-31411	Sequence 31411, A
65.4	17	1612	13	US-10-425-114-260	Sequence 260, App
65.4	17	1935	15	US-10-156-761-5202	Sequence 5202, App
65.4	17	42432	14	US-10-027-806-2	Sequence 2, Appl
65.4	17	42432	14	US-10-034-623-2	Sequence 2, Appl
65.4	17	42432	15	US-10-027-801-2	Sequence 2, Appl
65.4	17	42432	15	US-10-029-120-2	Sequence 2, Appl
65.4	17	93483	16	US-10-034-650-49	Sequence 49, Appl
65.4	17	9025608	15	US-10-156-761-1	Sequence 1, Appl
65.4	17	9025608	15	US-10-156-761-1	Sequence 1, Appl
65.4	17	1290	16	US-10-369-493-44453	Sequence 44453, A
64.6	16.8	1755	16	US-10-260-238-721	Sequence 721, App
64.6	16.8	5033	13	US-10-453-468-1	Sequence 1, Appl
63.8	16.6	290	9	US-09-294-093B-3144	Sequence 3144, Ap
63.8	16.6	864	16	US-10-260-238-3741	Sequence 3741, Ap
63.8	16.6	1284	16	US-10-369-493-35911	Sequence 35911, A
63.8	16.6	1407	16	US-10-438-784-4	Sequence 4, Appl
63.8	16.6	1455	13	US-10-282-122A-17996	Sequence 17996, A
63.8	16.6	2148	13	US-10-425-114-3454	Sequence 3454, Ap
63.8	16.6	2154	16	US-10-369-493-39263	Sequence 39263, A
63.8	16.6	2157	13	US-10-282-122A-15107	Sequence 15107, A
63.8	16.6	2160	16	US-10-369-493-39639	Sequence 39639, A
63.8	16.6	2160	16	US-10-369-493-40003	Sequence 40003, A
63.8	16.6	2466	16	US-10-369-493-42318	Sequence 42318, A
63.8	16.6	3807	16	US-10-438-784-2	Sequence 2, Appl
63.8	16.6	4218	15	US-10-017-161-1969	Sequence 1969, Ap
63.8	16.6	4218	16	US-10-292-798-1617	Sequence 1617, Ap
63.8	16.6	5898	16	US-10-438-784-1	Sequence 1, Appl
63.8	16.6	6077	15	US-10-311-455-1828	Sequence 1828, Ap
63.8	16.6	7543	10	US-09-993-731-3	Sequence 3, Appl
63.8	16.6	34769	13	US-10-221-714A-502	Sequence 502, App
63.8	16.6	67832	12	US-09-997-722-67	Sequence 67, Appl
63.1	16.4	266	13	US-10-424-599-3052	Sequence 3052, Ap
63.1	16.4	381	13	US-10-424-599-142107	Sequence 142107,
63.1	16.4	546	13	US-10-227-964C-1	Sequence 1, Appl
63.1	16.4	582	15	US-10-156-761-6486	Sequence 6486, Ap
63.1	16.4	627	16	US-10-369-493-41107	Sequence 41107, A
63.1	16.4	699	13	US-10-282-122A-30637	Sequence 30637, A
63.1	16.4	1422	13	US-10-282-122A-25899	Sequence 25899, A
63.1	16.4	1923	16	US-10-369-493-44696	Sequence 44696, A
63.1	16.4	2000	9	US-09-938-842A-4884	Sequence 4884, Ap
63.1	16.4	2000	11	US-09-938-842A-4884	Sequence 4884, Ap
63.1	16.4	2265	9	US-09-738-626-435	Sequence 435, App
63.1	16.4	2532	15	US-10-156-761-100	Sequence 100, App
63.1	16.4	2670	15	US-10-128-714-153	Sequence 153, App
63.1	16.4	2837	15	US-10-128-714-5153	Sequence 5153, Ap
63.1	16.4	6249	15	US-10-311-455-1188	Sequence 1188, Ap
63.1	16.4	17967	13	US-10-311-455-987	Sequence 987, App
63.1	16.4	46404	15	US-10-087-132-634	Sequence 634, App
63.1	16.4	330400	9	US-08-738-626-1	Sequence 1, Appl
62.3	16.2	214	13	US-10-085-733A-44852	Sequence 44852, A
62.3	16.2	214	16	US-10-242-535A-44852	Sequence 44852, A
62.3	16.2	330	11	US-09-864-408A-5843	Sequence 5843, Ap
62.3	16.2	371	9	US-09-738-626-2454	Sequence 2454, Ap
62.3	16.2	1185	15	US-10-156-761-6562	Sequence 6562, Ap
62.3	16.2	1353	13	US-10-282-122A-14430	Sequence 14430, A
62.3	16.2	1353	13	US-10-282-122A-11876	Sequence 11876, A
62.3	16.2	2073	13	US-10-282-122A-26442	Sequence 26442, A
62.3	16.2	2076	13	US-10-282-122A-28167	Sequence 28167, A
62.3	16.2	2748	9	US-09-738-626-2853	Sequence 2853, Ap
62.3	16.2	7256	17	US-10-239-090A-44	Sequence 44, Appl
62.3	16.2	8161	9	US-09-759-152-9	Sequence 9, Appl

88	16.2	62.3	8175	9	US-09-759-152-7	Sequence 7, Appli	161	15.6	60.0	460	13	US-10-085-783A-58370	Sequence 58370, A
c 89	16.2	62.3	150223	17	US-10-624-149A-1	Sequence 1, Appli	162	15.6	60.0	460	16	US-10-242-535A-58370	Sequence 58370, A
90	16.2	62.3	2940917	13	US-10-027-632-174763	Sequence 174763, A	163	15.6	60.0	500	13	US-10-085-783A-31184	Sequence 31184, A
91	16.2	62.3	2940917	16	US-10-027-632-174763	Sequence 174763, A	164	15.6	60.0	500	16	US-10-242-535A-31184	Sequence 31184, A
92	16.2	62.3	3309400	9	US-09-738-626-1	Sequence 1, Appli	165	15.6	60.0	531	15	US-10-156-761-5870	Sequence 5870, A
93	16	61.5	286	11	US-09-864-408A-3873	Sequence 3873, App	c 166	15.6	60.0	531	15	US-09-864-761-13268	Sequence 13268, A
94	16	61.5	363	15	US-10-156-761-721	Sequence 721, App	167	15.6	60.0	574	9	US-09-864-761-13696	Sequence 13696, A
95	16	61.5	385	13	US-10-424-599-134331	Sequence 134331, App	168	15.6	60.0	629	13	US-10-027-632-285790	Sequence 285790, A
96	16	61.5	387	9	US-09-867-701-1569	Sequence 1569, App	169	15.6	60.0	629	16	US-10-027-632-285790	Sequence 285790, A
97	16	61.5	406	10	US-09-803-719-2368	Sequence 2368, App	170	15.6	60.0	756	16	US-10-369-493-43105	Sequence 43105, A
98	16	61.5	441	15	US-10-156-761-1893	Sequence 1893, App	171	15.6	60.0	771	16	US-10-369-493-43105	Sequence 43105, A
99	16	61.5	489	9	US-09-894-844-63	Sequence 63, Appli	172	15.6	60.0	780	13	US-10-282-122A-29974	Sequence 29974, A
100	16	61.5	489	13	US-10-647-089-63	Sequence 63, Appli	173	15.6	60.0	832	13	US-10-221-625-207	Sequence 207, App
101	16	61.5	489	16	US-10-388-302-63	Sequence 63, Appli	c 174	15.6	60.0	870	15	US-10-156-761-3286	Sequence 3286, App
102	16	61.5	498	16	US-10-260-238-3297	Sequence 3297, App	175	15.6	60.0	870	15	US-10-260-238-3766	Sequence 3766, App
103	16	61.5	535	13	US-10-424-599-131246	Sequence 131246, A	176	15.6	60.0	960	15	US-10-156-761-1469	Sequence 1469, App
104	16	61.5	549	13	US-10-621-901-201	Sequence 201, App	177	15.6	60.0	1053	15	US-10-240-965-56	Sequence 56, Appli
c 105	16	61.5	604	13	US-10-027-632-275432	Sequence 275432, A	178	15.6	60.0	1075	15	US-10-219-220-216	Sequence 216, App
c 106	16	61.5	604	16	US-10-027-632-275432	Sequence 275432, A	179	15.6	60.0	1125	15	US-10-156-761-2293	Sequence 2293, App
107	16	61.5	779	13	US-10-027-632-154384	Sequence 154384, A	180	15.6	60.0	1131	16	US-10-369-493-33513	Sequence 33513, A
108	16	61.5	779	16	US-10-027-632-154384	Sequence 154384, A	c 181	15.6	60.0	1287	16	US-10-260-238-845	Sequence 845, App
109	16	61.5	941	10	US-09-991-936-1597	Sequence 1597, App	182	15.6	60.0	1389	15	US-10-230-026-27	Sequence 27, Appli
110	16	61.5	969	13	US-10-282-122A-13043	Sequence 13043, A	183	15.6	60.0	1413	16	US-10-369-493-35527	Sequence 35527, A
111	16	61.5	969	13	US-10-282-122A-14234	Sequence 14234, A	184	15.6	60.0	1443	15	US-10-156-761-3397	Sequence 3397, App
112	16	61.5	971	9	US-09-879-957-197	Sequence 197, App	185	15.6	60.0	1458	16	US-10-369-493-42392	Sequence 42392, A
113	16	61.5	988	16	US-10-198-706-1	Sequence 1, Appli	186	15.6	60.0	1470	13	US-10-282-122A-25960	Sequence 25960, A
c 114	16	61.5	1066	16	US-10-260-238-4037	Sequence 4037, App	187	15.6	60.0	1508	16	US-10-369-493-41474	Sequence 41474, A
115	16	61.5	1080	13	US-10-282-122A-33483	Sequence 33483, A	188	15.6	60.0	1508	13	US-10-282-122A-25889	Sequence 25889, A
116	16	61.5	1080	15	US-10-156-761-4054	Sequence 4054, App	189	15.6	60.0	2139	13	US-10-282-122A-11585	Sequence 11585, A
117	16	61.5	1116	15	US-10-156-761-6640	Sequence 6640, App	190	15.6	60.0	2160	13	US-10-430-011-76	Sequence 76, Appli
118	16	61.5	1135	13	US-10-424-599-33781	Sequence 33781, A	191	15.6	60.0	2160	13	US-10-430-011-83	Sequence 83, Appli
119	16	61.5	1167	16	US-10-369-493-34189	Sequence 34189, A	192	15.6	60.0	2160	13	US-10-430-011-126	Sequence 126, App
120	16	61.5	1188	15	US-10-156-761-6223	Sequence 6223, App	193	15.6	60.0	2160	15	US-10-138-927-76	Sequence 76, Appli
121	16	61.5	1197	15	US-10-156-761-5205	Sequence 5205, App	194	15.6	60.0	2160	15	US-10-138-927-83	Sequence 83, Appli
c 122	16	61.5	1227	9	US-09-974-300-1174	Sequence 1174, App	195	15.6	60.0	2291	13	US-10-282-122A-14324	Sequence 14324, A
123	16	61.5	1242	10	US-09-910-186A-21	Sequence 21, Appli	196	15.6	60.0	2337	13	US-10-282-122A-25485	Sequence 25485, A
124	16	61.5	1302	15	US-10-156-761-4554	Sequence 4554, App	c 197	15.6	60.0	2455	16	US-10-369-493-36725	Sequence 36725, A
125	16	61.5	1534	13	US-10-087-192-143	Sequence 143, App	198	15.6	60.0	2517	13	US-10-282-122A-25466	Sequence 25466, A
126	16	61.5	1557	16	US-10-369-493-44159	Sequence 44159, A	c 199	15.6	60.0	2547	13	US-10-282-122A-26047	Sequence 26047, A
127	16	61.5	1737	17	US-10-276-751-2	Sequence 2, Appli	200	15.6	60.0	2547	16	US-10-369-493-24476	Sequence 24476, A
128	16	61.5	1737	17	US-10-276-751-3	Sequence 3, Appli	c 201	15.6	60.0	2547	13	US-10-412-699B-1584	Sequence 1584, App
129	16	61.5	1872	13	US-10-425-114-15187	Sequence 15187, A	202	15.6	60.0	2849	15	US-10-156-761-361	Sequence 361, App
130	16	61.5	1881	16	US-10-369-493-28460	Sequence 28460, A	c 203	15.6	60.0	3750	13	US-10-282-122A-27453	Sequence 27453, A
131	16	61.5	1881	16	US-10-369-493-31420	Sequence 31420, A	204	15.6	60.0	10296	13	US-10-282-122A-33665	Sequence 33665, A
132	16	61.5	1884	17	US-10-276-751-1	Sequence 1, Appli	c 205	15.6	60.0	11050	15	US-10-204-708-85	Sequence 85, Appli
133	16	61.5	1932	16	US-10-369-493-41722	Sequence 41722, A	206	15.6	60.0	11275	15	US-10-240-689-39	Sequence 39, Appli
134	16	61.5	2034	15	US-10-156-761-3529	Sequence 3529, App	c 207	15.4	59.2	237	9	US-09-864-761-28053	Sequence 28053, A
135	16	61.5	2055	13	US-10-424-599-79230	Sequence 79230, A	208	15.4	59.2	253	13	US-10-424-599-151467	Sequence 151467, A
136	16	61.5	2205	16	US-10-424-599-79230	Sequence 103, App	209	15.4	59.2	271	13	US-10-424-599-86539	Sequence 86539, A
137	16	61.5	2364	14	US-10-282-122A-11362	Sequence 11362, A	210	15.4	59.2	275	11	US-09-923-876-278	Sequence 278, App
138	16	61.5	2559	14	US-10-011-588-24	Sequence 24, Appli	211	15.4	59.2	275	11	US-09-923-876-278	Sequence 278, App
139	16	61.5	2686	13	US-10-424-599-33779	Sequence 33779, A	212	15.4	59.2	293	13	US-10-424-599-10860	Sequence 10860, A
140	16	61.5	3318	13	US-10-282-122A-31347	Sequence 31347, A	213	15.4	59.2	293	13	US-10-029-386-15523	Sequence 15523, A
141	16	61.5	4079	16	US-10-305-720-1248	Sequence 1248, App	214	15.4	59.2	327	15	US-10-424-599-63343	Sequence 63343, A
142	16	61.5	4519	10	US-09-360-706-712	Sequence 712, App	215	15.4	59.2	369	13	US-10-027-632-289663	Sequence 289663, A
143	16	61.5	4519	17	US-10-641-643-1202	Sequence 1202, App	c 216	15.4	59.2	443	16	US-10-027-632-289663	Sequence 289663, A
144	16	61.5	4519	17	US-10-283-975A-223	Sequence 223, App	217	15.4	59.2	443	16	US-10-425-114-17731	Sequence 17731, A
145	16	61.5	4851	13	US-10-282-122A-25489	Sequence 25489, A	c 218	15.4	59.2	462	9	US-09-864-761-11462	Sequence 11462, A
146	16	61.5	88421	9	US-09-976-059-1	Sequence 1, Appli	219	15.4	59.2	463	13	US-10-424-599-125650	Sequence 125650, A
147	16	61.5	169276	13	US-10-087-192-142	Sequence 142, App	220	15.4	59.2	463	13	US-10-282-122A-23115	Sequence 23115, A
148	16	61.5	3186778	13	US-10-027-632-174961	Sequence 174961, A	221	15.4	59.2	477	13	US-10-424-599-76495	Sequence 76495, A
149	16	61.5	3186778	16	US-10-027-632-174961	Sequence 174961, A	222	15.4	59.2	480	15	US-10-156-761-3954	Sequence 3954, App
150	15.8	60.8	444	15	US-10-156-761-289	Sequence 289, App	223	15.4	59.2	488	16	US-10-260-238-111	Sequence 111, App
151	15.8	60.8	1119	16	US-10-369-493-28435	Sequence 28435, A	224	15.4	59.2	507	13	US-10-424-599-65821	Sequence 65821, A
152	15.8	60.8	1206	13	US-10-282-122A-13454	Sequence 13454, A	225	15.4	59.2	507	15	US-10-156-761-2692	Sequence 2692, App
153	15.8	60.8	2529	15	US-10-156-761-6449	Sequence 6449, App	226	15.4	59.2	520	13	US-10-424-599-125334	Sequence 125334, A
c 154	15.8	60.8	3865	13	US-10-087-192-964	Sequence 964, App	227	15.4	59.2	520	15	US-10-106-698-2652	Sequence 2652, App
155	15.6	60.0	134	9	US-09-864-761-29834	Sequence 29834, A	228	15.4	59.2	551	15	US-10-029-386-1823	Sequence 1823, App
c 156	15.6	60.0	333	16	US-10-085-783A-13248	Sequence 13248, A	229	15.4	59.2	571	10	US-03-918-995-6136	Sequence 6136, App
157	15.6	60.0	333	16	US-10-242-535A-13248	Sequence 13248, A	c 230	15.4	59.2	579	13	US-10-027-632-106025	Sequence 106025, A
c 158	15.6	60.0	382	9	US-09-783-590-6427	Sequence 6427, App	c 231	15.4	59.2	579	13	US-10-027-632-106026	Sequence 106026, A
159	15.6	60.0	441	15	US-10-156-761-5497	Sequence 5497, App	c 232	15.4	59.2	579	16	US-10-027-632-106025	Sequence 106025, A
160	15.6	60.0	456	15	US-10-219-220-14	Sequence 14, Appli	c 233	15.4	59.2	579	16	US-10-027-632-106026	Sequence 106026, A

Sequence 19986, A  
Sequence 19987, A  
Sequence 113615, A  
Sequence 113615, A  
Sequence 113615, A  
Sequence 175089, A  
Sequence 175089, A  
Sequence 13233, A  
Sequence 4, Appl1  
Sequence 40038, A  
Sequence 37104, A  
Sequence 15269, A  
Sequence 7921, Ap  
Sequence 3552, Ap  
Sequence 3958, Ap  
Sequence 2049, Ap  
Sequence 2612, Ap  
Sequence 1612, Ap  
Sequence 16268, A  
Sequence 11940, A  
Sequence 13179, A  
Sequence 4455, Ap  
Sequence 738, Ap  
Sequence 2079, Ap  
Sequence 28095, A  
Sequence 14443, A  
Sequence 71, Appl  
Sequence 71, Appl  
Sequence 71, Appl  
Sequence 71, Appl  
Sequence 40487, A  
Sequence 32573, A  
Sequence 6808, Ap  
Sequence 6056, Ap  
Sequence 9, Appl  
Sequence 375, Ap  
Sequence 91, Appl  
Sequence 375, Ap  
Sequence 22478, A  
Sequence 1049, Ap  
Sequence 14737, A  
Sequence 30, Appl  
Sequence 28056, A  
Sequence 45407, A  
Sequence 23501, A  
Sequence 5, Appl  
Sequence 6, Appl  
Sequence 31004, A  
Sequence 110126, A  
Sequence 43077, A  
Sequence 13208, A  
Sequence 23265, A  
Sequence 35492, A  
Sequence 13623, A  
Sequence 44721, A  
Sequence 17493, A  
Sequence 42285, A  
Sequence 17722, A  
Sequence 17876, A  
Sequence 19656, A  
Sequence 174, App  
Sequence 27853, A  
Sequence 32155, A

US-10-027-632-19986  
US-10-027-632-19987  
US-10-027-632-113615  
US-10-027-632-113615  
US-10-027-632-113615  
US-10-027-632-175089  
US-10-027-632-175089  
US-10-282-122A-13233  
US-10-425-114-4  
US-10-282-122A-40038  
US-10-282-122A-37104  
US-10-425-114-15269  
US-09-815-242-7921  
US-10-424-599-3552  
US-10-425-114-3958  
US-10-320-797-2049  
US-10-156-761-2612  
US-10-425-114-16268  
US-10-282-122A-11940  
US-10-425-114-13179  
US-10-156-761-4455  
US-10-259-165-738  
US-10-425-114-2079  
US-10-425-114-28095  
US-10-425-114-14443  
US-10-027-866-71  
US-10-034-623-71  
US-10-027-801-71  
US-10-029-120-71  
US-10-369-493-40487  
US-10-425-114-32573  
US-10-156-761-6808  
US-10-156-761-6056  
US-10-440-503-91  
US-10-440-503-375  
US-10-461-925-91  
US-10-461-925-375  
US-10-146-772-91  
US-10-146-772-375  
US-10-241-742-91  
US-10-241-742-375  
US-10-440-523-91  
US-10-440-523-375  
US-10-425-114-22478  
US-10-320-797-1049  
US-10-282-122A-14737  
US-10-264-237-30  
US-10-282-122A-28056  
US-10-369-493-45407  
US-10-282-122A-23501  
US-10-460-138-5  
US-10-460-138-6  
US-10-369-493-31004  
US-10-424-599-110126  
US-10-369-493-43077  
US-10-282-122A-13208  
US-10-425-114-23265  
US-10-425-114-35492  
US-10-282-122A-13623  
US-10-369-493-44721  
US-10-425-114-17493  
US-10-369-493-42285  
US-10-425-114-17722  
US-10-425-114-17876  
US-10-282-122A-19656  
US-10-313-542-174  
US-10-425-114-27853  
US-10-282-122A-32155

ALIGNMENTS

Query Match 100.0%; Score 26; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.0076;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-09-996-634-75/c  
; Sequence 75, Application US/09996634  
; Patent No. US20020172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 61260  
; CURRENT APPLICATION NUMBER: US/09/996,634  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/447,135  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08/990,823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 75  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(401)  
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.  
; US-09-996-634-75

Query Match 100.0%; Score 26; DB 9; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.0076;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCGCGAATCTACGAA 26  
DB 148 ACAACGGCGAGCGCGAATCTACGAA 123

RESULT 2  
US-09-997-182-75/c  
; Sequence 75, Application US/09997182  
; Publication No. US20030049263A1  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 61258  
; CURRENT APPLICATION NUMBER: US/09/997,182  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/447,135  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08/990,823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 75  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(401)  
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.  
; US-09-997-182-75

ALIGNMENTS

Query Match 100.0%; Score 26; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 0.0076;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 67461
; LENGTH: 806
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_31931C.1
US-10-424-599-67461

Query Match      68.58; Score 17.8; DB 13; Length 806;
Best Local Similarity 90.58; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      5 CCGCGAGCGCCCGAATCTACGA 25
Db      444 CCGCGAGCGCCCGGATCAACGA 464

RESULT 7
US-10-029-386-6930
; Sequence 6930, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AFOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6930
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009336.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: SWISSPROT HIT: Q03828, EVALUE 3.00e-18
; OTHER INFORMATION: EST HUMAN HIT: BG489979.1, EVALUE 4.80e-02
; OTHER INFORMATION: NT HIT: S34322.1, EVALUE 1.00e-41
US-10-029-386-6930

Query Match      67.78; Score 17.6; DB 15; Length 531;
Best Local Similarity 83.38; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 AACCGCGAGCGCCCGAATCTACGAA 26
Db      506 AACCGCGTGGCCCGGACGACGGA 529

RESULT 8
US-10-425-114-22681
; Sequence 22681, Application US/10425114
; Publication No. US20040034889A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
```

```
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22681
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-095-F7_FLI
US-10-425-114-22681

Query Match      67.78; Score 17.6; DB 13; Length 612;
Best Local Similarity 83.38; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 CAACGCGAGCGCCCGAATCTACGA 25
Db      174 CAACGCGCGGCTACATCTACGA 197

RESULT 9
US-10-282-122A-12032
; Sequence 12032, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA_034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12032
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
```

```
US-10-282-122A-12032
Query Match      67.7%; Score 17.6; DB 13; Length 1017;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGCAATCTACGA 25
  |||||
Db 798 CATCGCGAGCGCCCACTACGA 821

RESULT 10
US-10-369-493-44269
; Sequence 44269, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44269
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Rhodopsseudomonas palustris
US-10-369-493-44269

Query Match      67.7%; Score 17.6; DB 16; Length 2205;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGGCCCGCAATCTACGA 26
  |||||
Db 1996 AACGGCGAGGTCGAGATCTACGA 2019

RESULT 11
US-09-939-964-1/c
; Sequence 1, Application US/09939964
; Publication No. US20030054522A1
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; FILE REFERENCE: Plasmid
; CURRENT APPLICATION NUMBER: US/09/939,964
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/214,808
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-939-964-1

Query Match      67.7%; Score 17.6; DB 10; Length 536165;
Best Local Similarity 83.3%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGGCCCGCAATCTACGA 26
  |||||
Db 1996 AACGGCGAGGTCGAGATCTACGA 2019

US-10-282-122A-12032
Query Match      67.7%; Score 17.6; DB 13; Length 1017;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGCAATCTACGA 25
  |||||
Db 798 CATCGCGAGCGCCCACTACGA 821

RESULT 10
US-10-369-493-44269
; Sequence 44269, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44269
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Rhodopsseudomonas palustris
US-10-369-493-44269

Query Match      67.7%; Score 17.6; DB 16; Length 2205;
Best Local Similarity 83.3%; Pred. No. 87;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGGCCCGCAATCTACGA 26
  |||||
Db 1996 AACGGCGAGGTCGAGATCTACGA 2019

RESULT 11
US-09-939-964-1/c
; Sequence 1, Application US/09939964
; Publication No. US20030054522A1
; GENERAL INFORMATION:
; APPLICANT: Rosenthal, Andre
; APPLICANT: Freiberg, Christoph
; APPLICANT: Perret, Xavier Philippe
; APPLICANT: Broughton, William John
; TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
; FILE REFERENCE: Plasmid
; CURRENT APPLICATION NUMBER: US/09/939,964
; PRIOR FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/214,808
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 536165
; TYPE: DNA
; ORGANISM: Rhizobium
US-09-939-964-1

Query Match      67.7%; Score 17.6; DB 10; Length 536165;
Best Local Similarity 83.3%; Pred. No. 84;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AACGGCGAGGCCCGCAATCTACGA 26
  |||||
Db 1996 AACGGCGAGGTCGAGATCTACGA 2019

US-10-624-599-110822
Query Match      65.4%; Score 17; DB 13; Length 383;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
  |||||
Db 315 ACAACGCTAGGCGCTGAATTCGCA 339

RESULT 13
US-10-156-761-2378
; Sequence 2378, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2378
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(669)
US-10-156-761-2378

Query Match      65.4%; Score 17; DB 15; Length 669;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
  |||||
Db 278 AGAACGTCGAGCGCTGATCTACGA 302

US-10-624-599-110822
Query Match      65.4%; Score 17; DB 13; Length 383;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
  |||||
Db 315 ACAACGCTAGGCGCTGAATTCGCA 339

RESULT 13
US-10-156-761-2378
; Sequence 2378, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2378
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(669)
US-10-156-761-2378

Query Match      65.4%; Score 17; DB 15; Length 669;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGCAATCTACGA 25
  |||||
Db 278 AGAACGTCGAGCGCTGATCTACGA 302
```



## RESULT 14

US-10-369-493-27924  
; Sequence 27924, Application US/10369493  
; Publication No. US2002033675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 27924  
; LENGTH: 672  
; TYPE: DNA  
; ORGANISM: Neurospora crassa  
US-10-369-493-27924

Query Match 65.4%; Score 17; DB 16; Length 672;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGATCTACGAA 26

DB 303 CAACACGAGACCGCATCTACGAA 327

## RESULT 15

US-10-027-632-168321/c  
; Sequence 168321, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 168321  
; LENGTH: 821  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-168321

Query Match 65.4%; Score 17; DB 13; Length 821;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCGCCGATCTACGAA 26

DB 509 CAACCCCGAGCGCATGAATCTACCAA 485

## RESULT 16

US-10-027-632-168322/c  
; Sequence 168322, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 168322  
; LENGTH: 821  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-168322

Query Match 65.4%; Score 17; DB 13; Length 821;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACCGCGAGCGCCGATCTACGAA 26

DB 509 CAACCCCGAGCGCATGAATCTACCAA 485

## RESULT 17

US-10-027-632-168321/c  
; Sequence 168321, Application US/10027632  
; Publication No. US20020204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
US-10-027-632-168321

; SEQ ID NO 168321  
; LENGTH: 821  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-168321

Query Match 65.4%; Score 17; DB 16; Length 821;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGAATCTACGAA 26  
||| ||||| ||||| ||||| |||||  
Db 509 CAACCCGAGGCATGAATCTACCAA 485

## RESULT 18

US-10-027-632-168322/c  
; Sequence 168322, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 168322  
; LENGTH: 821  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-168322

Query Match 65.4%; Score 17; DB 16; Length 821;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGCCCGAATCTACGAA 26  
||| ||||| ||||| ||||| |||||  
Db 509 CAACCCGAGGCATGAATCTACCAA 485

## RESULT 19

US-10-027-806-37/c  
; Sequence 37, Application US/10027806  
; Publication No. US20020160476A1  
; GENERAL INFORMATION:

; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCCRP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,806  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 37  
; LENGTH: 879  
; TYPE: DNA  
; ORGANISM: Cenarchaeum symbiosum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(879)  
US-10-027-806-37

Query Match 65.4%; Score 17; DB 14; Length 879;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGAATCTACGA 25  
||| ||||| ||||| ||||| |||||  
Db 434 ACCTCGAGAGGCCCGAGTCTACAA 410

## RESULT 20

US-10-034-623-37/c

; Sequence 37, Application US/10034623  
; Publication No. US20020198365A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCCRP.002A  
; CURRENT APPLICATION NUMBER: US/10/034,623  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: 60/102,294  
; PRIOR FILING DATE: 1998-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37  
; LENGTH: 879  
; TYPE: DNA  
; ORGANISM: Cenarchaeum symbiosum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(879)  
US-10-034-623-37

Query Match 65.4%; Score 17; DB 14; Length 879;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGCCCGAATCTACGA 25  
||| ||||| ||||| ||||| |||||  
Db 434 ACCTCGAGAGGCCCGAGTCTACAA 410

## RESULT 21

US-10-027-801-37/c  
; Sequence 37, Application US/10027801  
; Publication No. US20030054364A1  
; GENERAL INFORMATION:

; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCCRP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,801  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020  
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 37  
; LENGTH: 879  
; TYPE: DNA

```

; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(879)
US-10-027-801-37

Query Match          65.4%; Score 17; DB 15; Length 879;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCCGGAATCTACGA 25
Db 434 ACCTCGGAGAGGCCCGGAGTCTACAA 410

RESULT 22
US-10-029-120-37/c
; Sequence 37, Application US/10029120
; Publication No. US20030175708A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOOP.002A
; CURRENT APPLICATION NUMBER: US/10/029,120
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(879)
US-10-029-120-37

Query Match          65.4%; Score 17; DB 15; Length 879;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCCGGAATCTACGA 25
Db 434 ACCTCGGAGAGGCCCGGAGTCTACAA 410

RESULT 23
US-10-282-122A-13773/c
; Sequence 13773, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13773

Query Match          65.4%; Score 17; DB 13; Length 1173;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CAACGGCGAGGCCCGGAATCTACGAA 26
Db 199 CGACGGCGAGGCCCGCGCTTCGAA 175

RESULT 24
US-10-425-114-23453/c
; Sequence 23453, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 23453
; LENGTH: 1391
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3596-038-G7_FLI
US-10-425-114-23453

Query Match          65.4%; Score 17; DB 13; Length 1391;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCCGGAATCTACGA 25
Db 392 ACACGGCGGTGGCGCTGCATCGACA 368

RESULT 25
US-10-369-493-28653
; Sequence 28653, Application US/10369493
; Publication No. US20030233675A1
```

; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 28653  
; LENGTH: 1557  
; TYPE: DNA  
; ORGANISM: Burkholderia fungorum  
US-10-369-493-28653

Query Match 65.4%; Score 17; DB 16; Length 1557;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGAATCTACGAA 26  
DB 807 CGACGGCGAGTTCGCGATGTACGAA 831

## RESULT 26

US-10-369-493-31411  
; Sequence 31411, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 31411  
; LENGTH: 1563  
; TYPE: DNA  
; ORGANISM: Burkholderia cepacia  
US-10-369-493-31411

Query Match 65.4%; Score 17; DB 16; Length 1563;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGGCGAGGCCCGAATCTACGAA 26  
DB 813 CGACGGCGAGTTCGCGATGTACGAA 837

## RESULT 27

US-10-425-114-260  
; Sequence 260, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 260  
; LENGTH: 1612  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700051638\_FLI  
US-10-425-114-260

Query Match 65.4%; Score 17; DB 13; Length 1612;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGAATCTACGA 25  
DB 79 ACACGGCGAGGCCCGAGGCTCTCGGA 103

## RESULT 28

US-10-156-761-5202  
; Sequence 5202, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAVOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5202  
; LENGTH: 1935  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1935)  
US-10-156-761-5202

Query Match 65.4%; Score 17; DB 15; Length 1935;  
Best Local Similarity 80.0%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ACAACGGCGAGGCCCGAATCTACGA 25  
DB 1424 ACATCTCCAGGCCCGGATCTACGA 1448

## RESULT 29

US-10-027-806-2/c  
; Sequence 2, Application US/10027806  
; Publication No. US20020160476A1  
; GENERAL INFORMATION:  
; APPLICANT: Swanson, Ronald V.  
; APPLICANT: Feldman, Robert A.  
; APPLICANT: Schleper, Christa  
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM  
; FILE REFERENCE: DCOEP.002A  
; CURRENT APPLICATION NUMBER: US/10/027,806

```

;
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; NAME/KEY: CDS
; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
; US-10-027-806-2

Query Match          65.4%; Score 17; DB 14; Length 42432;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 37837 ACCTCGGAGAGCGCGGAGTCTACAA 37813

Search completed: June 20, 2004, 17:46:13
Job time : 162.058 secs

```

```

;
; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
; US-10-034-623-2

Query Match          65.4%; Score 17; DB 14; Length 42432;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ACAACGGCGAGGCCGGAATCTACGA 25
Db 37837 ACCTCGGAGAGCGCGGAGTCTACAA 37813

Search completed: June 20, 2004, 17:46:13
Job time : 162.058 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 109.175 Seconds  
(without alignments)  
739.327 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19

Sequence: 1 gtgcagcgcgcgatgacc 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 0%

Maximum DB seq length: 2000000000

Listing first 300 summaries

Database : N Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001as.\*

5: Geneseq2001bs.\*

6: Geneseq2002s.\*

7: Geneseq2003as.\*

8: Geneseq2003bs.\*

9: Geneseq2003cs.\*

10: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	393	2	AA19174 Partial D
2	19	100.0	110000	4	AA199682_42 Continuation (43 o
3	19	100.0	110000	4	AA199683_42 Continuation (43 o
4	17	89.5	25	8	AC137969 Human mic
5	16.4	86.3	270	6	AB172618 Corn tass
6	16.4	86.3	432	3	AA139698 Zea tass
7	16.4	86.3	445	3	AA139698 Zea tass
8	16.4	86.3	963	7	ADA70332 Rice gene
9	16.4	86.3	1077	8	ADA8367 Rice gene
10	16.4	86.3	1560	6	AA199683_42 Continuation (43 o
11	16.4	86.3	1948	7	AA199683_42 Continuation (43 o
12	16.4	86.3	2000	6	AA199683_42 Continuation (43 o
13	16.4	86.3	47475	3	AA199683_42 Continuation (43 o
14	16.4	86.3	110000	3	AA199683_42 Continuation (43 o
15	16.4	86.3	110000	4	AA199683_42 Continuation (43 o
16	16.4	86.3	110000	4	AA199683_42 Continuation (43 o
17	16.4	86.3	34980	3	AA199683_42 Continuation (43 o
18	16.4	86.3	34980	6	AA199683_42 Continuation (43 o
19	16	84.2	942	9	AA199683_42 Continuation (43 o
20	16	84.2	1821	3	AA199683_42 Continuation (43 o
21	15.8	83.2	51	4	AA199683_42 Continuation (43 o
22	15.8	83.2	51	4	AA199683_42 Continuation (43 o
23	15.8	83.2	51	4	AA199683_42 Continuation (43 o

7	ACA42539	477	15.8	83.2	3	AA199683_42 Continuation (43 o
8	AA199683_42	598	15.8	83.2	3	AA199683_42 Continuation (43 o
9	AA199683_42	598	15.8	83.2	3	AA199683_42 Continuation (43 o
10	AA199683_42	609	15.8	83.2	3	AA199683_42 Continuation (43 o
11	AA199683_42	620	15.8	83.2	3	AA199683_42 Continuation (43 o
12	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
13	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
14	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
15	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
16	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
17	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
18	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
19	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
20	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
21	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
22	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
23	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
24	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
25	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
26	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
27	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
28	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
29	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
30	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
31	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
32	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
33	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
34	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
35	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
36	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
37	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
38	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
39	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
40	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
41	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
42	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
43	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
44	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
45	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
46	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
47	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
48	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
49	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
50	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
51	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
52	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
53	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
54	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
55	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
56	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
57	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
58	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
59	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
60	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
61	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
62	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
63	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
64	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
65	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
66	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
67	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
68	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
69	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
70	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
71	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
72	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
73	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
74	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
75	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
76	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
77	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
78	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
79	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
80	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
81	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
82	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
83	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
84	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
85	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
86	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
87	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
88	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
89	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
90	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
91	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
92	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
93	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
94	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
95	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o
96	AA199683_42	738	15.8	83.2	3	AA199683_42 Continuation (43 o

97	15.4	81.1	530	6	ABQ17185	Abq17185 Oligonucle	C 170	14.8	77.9	120	6	ABL75421	Ab175421 Corn tass
98	15.4	81.1	530	6	ABQ17184	Abq17184 Oligonucle	C 171	14.8	77.9	221	6	ABN18064	Abn18064 Human ORF
99	15.4	81.1	610	7	ABX96466	ABX96466 Rice endo	C 172	14.8	77.9	234	9	Add16032	Add16032 CDNA (Seq
100	15.4	81.1	729	6	ABQ26106	ABQ26106 Oligonucle	C 173	14.8	77.9	271	7	ABX84396	ABX84396 Corn ear
101	15.4	81.1	729	6	ABQ26107	ABQ26107 Oligonucle	C 174	14.8	77.9	282	7	ABX87965	ABX87965 Corn ear
102	15.4	81.1	732	7	ACB25151	ACB25151 Prokaryot	C 175	14.8	77.9	298	7	ABX87583	ABX87583 Corn ear
103	15.4	81.1	735	7	ACA38066	ACA38066 Prokaryot	C 176	14.8	77.9	315	3	AAA31887	AAA31887 Plant mic
104	15.4	81.1	825	7	ACA45245	ACA45245 Prokaryot	C 177	14.8	77.9	318	6	ABN76240	ABN76240 Human ORF
105	15.4	81.1	851	6	ABQ29083	ABQ29083 Oligonucle	C 178	14.8	77.9	354	8	ACL11875	ACL11875 DNA clone
106	15.4	81.1	851	6	ABQ29082	ABQ29082 Oligonucle	C 179	14.8	77.9	355	2	AAA34158	AAA34158 Mycobacte
107	15.4	81.1	903	6	ABQ14805	ABQ14805 Oligonucle	C 180	14.8	77.9	356	2	AAA34157	AAA34157 Mycobacte
108	15.4	81.1	903	6	ABQ14804	ABQ14804 Oligonucle	C 181	14.8	77.9	357	2	AAA34156	AAA34156 Mycobacte
109	15.4	81.1	945	7	ACA35953	ACA35953 Prokaryot	C 182	14.8	77.9	392	4	AAK56036	AAK56036 Human inm
110	15.4	81.1	981	7	ACA49565	ACA49565 Prokaryot	C 183	14.8	77.9	438	6	ABN17454	ABN17454 Human ORF
111	15.4	81.1	999	7	ACA50943	ACA50943 Prokaryot	C 184	14.8	77.9	444	7	ABQ84702	ABQ84702 Manduca s
112	15.4	81.1	1010	6	ABQ28481	ABQ28481 Oligonucle	C 185	14.8	77.9	447	7	ACA41903	ACA41903 Prokaryot
113	15.4	81.1	1010	6	ABQ28480	ABQ28480 Oligonucle	C 186	14.8	77.9	468	7	ABZ38188	ABZ38188 N. gonorr
114	15.4	81.1	1035	7	ACA51449	ACA51449 Prokaryot	C 187	14.8	77.9	468	7	ACA23531	ACA23531 Prokaryot
115	15.4	81.1	1035	7	ACA51448	ACA51448 Pseudomon	C 188	14.8	77.9	468	7	ACA40957	ACA40957 Prokaryot
116	15.4	81.1	1176	7	ACA51486	ACA51486 Prokaryot	C 189	14.8	77.9	468	7	ACA26504	ACA26504 Prokaryot
117	15.4	81.1	1197	4	AAH51360	AAH51360 Mycobacte	C 190	14.8	77.9	471	7	ACA41649	ACA41649 Prokaryot
118	15.4	81.1	1265	6	ABK75033	ABK75033 Bacillus	C 191	14.8	77.9	480	5	AAH88727	AAH88727 Cellulose
119	15.4	81.1	1361	6	ABK75032	ABK75032 Drosophil	C 192	14.8	77.9	480	8	ACL11879	ACL11879 DNA clone
120	15.4	81.1	1515	7	ACA38048	ACA38048 Prokaryot	C 193	14.8	77.9	493	3	AAH30710	AAH30710 Human col
121	15.4	81.1	1576	4	ABL10421	ABL10421 Drosophil	C 194	14.8	77.9	493	3	AAH30710	AAH30710 Human col
122	15.4	81.1	1599	4	AAH71430	AAH71430 Corynebac	C 195	14.8	77.9	500	6	ABQ34051	ABQ34051 Oligonucle
123	15.4	81.1	1713	5	AAH65969	AAH65969 C. Glutami	C 196	14.8	77.9	500	6	ABQ34050	ABQ34050 Oligonucle
124	15.4	81.1	1836	4	AAH71428	AAH71428 Corynebac	C 197	14.8	77.9	501	6	ABQ28996	ABQ28996 Oligonucle
125	15.4	81.1	1923	6	ABZ14337	ABZ14337 Arabidops	C 198	14.8	77.9	501	6	ABQ28997	ABQ28997 Oligonucle
126	15.4	81.1	1928	2	AAH95762	AAH95762 Arabidops	C 199	14.8	77.9	523	8	ACL11933	ACL11933 DNA clone
127	15.4	81.1	1928	3	AAH65290	AAH65290 Arabidops	C 200	14.8	77.9	534	7	ABZ39636	ABZ39636 N. gonorr
128	15.4	81.1	1928	3	AAH65290	AAH65290 Arabidops	C 201	14.8	77.9	542	8	ACL11947	ACL11947 DNA clone
129	15.4	81.1	2110	3	AAH65290	AAH65290 Arabidops	C 202	14.8	77.9	551	8	ACL11892	ACL11892 DNA clone
130	15.4	81.1	2602	4	ABZ126431	ABZ126431 Drosophil	C 203	14.8	77.9	553	8	ACL11905	ACL11905 DNA clone
131	15.4	81.1	2832	4	ABZ126431	ABZ126431 Drosophil	C 204	14.8	77.9	554	8	ACL11905	ACL11905 DNA clone
132	15.4	81.1	2832	4	ABZ126431	ABZ126431 Drosophil	C 205	14.8	77.9	555	8	ABZ39635	ABZ39635 N. gonorr
133	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 206	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
134	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 207	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
135	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 208	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
136	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 209	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
137	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 210	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
138	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 211	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
139	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 212	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
140	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 213	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
141	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 214	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
142	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 215	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
143	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 216	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
144	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 217	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
145	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 218	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
146	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 219	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
147	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 220	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
148	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 221	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
149	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 222	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
150	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 223	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
151	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 224	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
152	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 225	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
153	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 226	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
154	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 227	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
155	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 228	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
156	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 229	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
157	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 230	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
158	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 231	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
159	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 232	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
160	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 233	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
161	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 234	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
162	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 235	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
163	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 236	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
164	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 237	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
165	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 238	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
166	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 239	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
167	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 240	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
168	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 241	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone
169	15.4	81.1	3306	2	AAQ22202	AAQ22202 A. chryso	C 242	14.8	77.9	558	8	ACL11932	ACL11932 DNA clone



C 243 14.8 77.9 632 8 ACL11867  
C 244 14.8 77.9 632 8 ACL11931  
C 245 14.8 77.9 632 8 ACL11914  
C 246 14.8 77.9 632 8 ACL11914  
C 247 14.8 77.9 632 8 ACL11925  
C 248 14.8 77.9 636 8 ACL11866  
C 249 14.8 77.9 639 8 ACL11877  
C 250 14.8 77.9 642 8 ACL11951  
C 251 14.8 77.9 647 8 ACL11926  
C 252 14.8 77.9 651 8 ACL11873  
C 253 14.8 77.9 651 8 ACL11878  
C 254 14.8 77.9 656 8 ACL11921  
C 255 14.8 77.9 657 8 ACL11927  
C 256 14.8 77.9 664 8 ACL11919  
C 257 14.8 77.9 665 8 ACL11938  
C 258 14.8 77.9 668 7 ABZ53865  
C 259 14.8 77.9 674 8 ACL11936  
C 260 14.8 77.9 678 8 ACL11916  
C 261 14.8 77.9 681 8 ACL11865  
C 262 14.8 77.9 688 8 ACL11896  
C 263 14.8 77.9 694 8 ACL11891  
C 264 14.8 77.9 698 8 ACL11871  
C 265 14.8 77.9 700 4 RAH92000  
C 266 14.8 77.9 720 8 ACL11868  
C 267 14.8 77.9 726 8 ADA48757  
C 268 14.8 77.9 735 7 ACA43973  
C 269 14.8 77.9 741 4 AAS54206  
C 270 14.8 77.9 741 7 ACA42468  
C 271 14.8 77.9 741 7 ACA40915  
C 272 14.8 77.9 787 3 AAS58916  
C 273 14.8 77.9 789 5 AAH65192  
C 274 14.8 77.9 789 7 ACA01924  
C 275 14.8 77.9 818 8 ADA48139  
C 276 14.8 77.9 819 7 ACF39432  
C 277 14.8 77.9 822 7 ACA38666  
C 278 14.8 77.9 822 7 ACA40868  
C 279 14.8 77.9 825 2 AAQ11855  
C 280 14.8 77.9 850 3 AAF14891  
C 281 14.8 77.9 854 7 ADA69341  
C 282 14.8 77.9 874 6 ABQ18184  
C 283 14.8 77.9 874 6 ABQ18185  
C 284 14.8 77.9 879 7 ACA23319  
C 285 14.8 77.9 885 6 ABK73038  
C 286 14.8 77.9 902 6 ABQ21053  
C 287 14.8 77.9 902 6 ABQ21052  
C 288 14.8 77.9 918 6 ABQ90033  
C 289 14.8 77.9 918 6 ABQ90034  
C 290 14.8 77.9 933 4 AAH43317  
C 291 14.8 77.9 933 9 AAD59980  
C 292 14.8 77.9 954 6 ABA94410  
C 293 14.8 77.9 966 6 ABH89590  
C 294 14.8 77.9 966 7 ACA42108  
C 295 14.8 77.9 972 7 ACA37709  
C 296 14.8 77.9 987 5 AAH68462  
C 297 14.8 77.9 993 9 ADC23846  
C 298 14.8 77.9 1005 3 AAF14888  
C 299 14.8 77.9 1009 5 AAH8784  
C 300 14.8 77.9 1012 6 ABK73256

## ALIGNMENTS

RESULT 1  
AAT49174  
ID AAT49174 standard; DNA; 393 BP.  
XX  
AC AAT49174;  
XX  
DT 16-OCT-1997 (first entry)  
XX  
DE Partial DNA clone HinP1-27 encoding immunostimulatory peptide.  
XX

KW Mycobacterium tuberculosis; immunostimulatory peptide; T cell epitope;  
KW stimulate; interferon-gamma; IFN-gamma; production; vaccine;  
KW tuberculin skin test; ds.  
OS Mycobacterium tuberculosis.  
PN WO9700067-A1.  
PD 03-JAN-1997.  
XX  
XX 14-JUN-1996; 96WO-US010375.  
XX  
PR 15-JUN-1995; 95US-0000254P.  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
PI Nano FE;  
DR WPI; 1997-077347/07.  
XX  
XX New immuno-stimulatory peptide(s) of Mycobacterium tuberculosis - useful  
PT in vaccines, diagnostic skin test, immunoassay and gene isolation.  
XX  
PS Claim 1; Page 1; 79pp; English.  
XX  
CC AAT49100-175 are DNA sequences isolated from Mycobacterium tuberculosis,  
CC which encode partial sequences of immunostimulatory peptides. Each of the  
CC clones encode at least one immunostimulatory T cell epitope. The clones  
CC were identified by testing over 300 fusion clones (alkaline phosphatase-  
CC M. tuberculosis peptide fusions) for their ability to stimulate  
CC interferon (IFN)-gamma production. 80 clones were initially designated to  
CC have some ability to stimulate IFN-gamma production, of which 76 are  
CC shown in AAT49100-175. These sequences can be used to obtain the full  
CC length M. tuberculosis genes and corresponding proteins using standard  
CC techniques. The peptides are useful in vaccines, as reagents in an  
CC improved tuberculin skin test (especially using peptides different from  
CC those used in vaccines so as to allow differentiation between vaccinated  
CC and infected subjects), and as immunoassay reagents for detecting specific  
CC antibodies. An advantage of these peptides is that they stimulate  
CC production of IFN-gamma (critical for a protective immune response to M.  
CC tuberculosis) by CD4-positive T cells  
SQ Sequence 393 BP; 72 A; 120 C; 119 G; 82 T; 0 U; 0 Other;

Query Match 100.0%; Score 19; DB 2; Length 393;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCCGCATGACC 19  
Db 12 GTCGACGCCCGCATGACC 30

## RESULT 2

AAT49174  
Continuation (43 of 45) of AAT49174 from base 4200001 (Mycobacterium tuberculosis strain  
WP Sequence split into 45 fragments LOCUS AAT49174 Accession AAT49174  
WP Fragment Name Begin End  
WP AAT49174\_01 1 110000  
WP AAT49174\_02 100001 210000  
WP AAT49174\_03 200001 310000  
WP AAT49174\_04 300001 410000  
WP AAT49174\_05 400001 510000  
WP AAT49174\_06 500001 610000  
WP AAT49174\_07 600001 710000  
WP AAT49174\_08 700001 810000  
WP AAT49174\_09 800001 910000  
WP AAT49174\_10 900001 1010000  
WP AAT49174\_11 100001 1110000  
WP AAT49174\_12 110001 1210000  
WP AAT49174\_13 120001 1310000  
WP AAT49174\_14 130001 1410000  
WP AAT49174\_15 140001 1510000

WP	AAI9682_15	1500001	1610000
WP	AAI9682_16	1600001	1710000
WP	AAI9682_17	1700001	1810000
WP	AAI9682_18	1800001	1910000
WP	AAI9682_19	1900001	2010000
WP	AAI9682_20	2000001	2110000
WP	AAI9682_21	2100001	2210000
WP	AAI9682_22	2200001	2310000
WP	AAI9682_23	2300001	2410000
WP	AAI9682_24	2400001	2510000
WP	AAI9682_25	2500001	2610000
WP	AAI9682_26	2600001	2710000
WP	AAI9682_27	2700001	2810000
WP	AAI9682_28	2800001	2910000
WP	AAI9682_29	2900001	3010000
WP	AAI9682_30	3000001	3110000
WP	AAI9682_31	3100001	3210000
WP	AAI9682_32	3200001	3310000
WP	AAI9682_33	3300001	3410000
WP	AAI9682_34	3400001	3510000
WP	AAI9682_35	3500001	3610000
WP	AAI9682_36	3600001	3710000
WP	AAI9682_37	3700001	3810000
WP	AAI9682_38	3800001	3910000
WP	AAI9682_39	3900001	4010000
WP	AAI9682_40	4000001	4110000
WP	AAI9682_41	4100001	4210000
WP	AAI9682_42	4200001	4310000
WP	AAI9682_43	4300001	4410000
WP	AAI9682_44	4400001	4411529

Query Match  
Best Local Similarity 100.0%;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Score 19; DB 4; Length 110000;  
Pred. No. 38;

QY

1 GTGACGCCGGCGATGACC 19  
|||||

DB

27060 GTGACGCCGGCGATGACC 27042  
|||||

RESULT 3

AAI9683\_42/c  
Continuation (43 of 44) of AAI9683 from base 4200001 (Mycobacterium tuberculosis strain  
WP Sequence split into 44 fragments LOCUS AAI9683 Accession Aai9683  
Begin End  
Fragment Name  
WP AAI9683\_00 1 110000  
WP AAI9683\_01 100001 210000  
WP AAI9683\_02 200001 310000  
WP AAI9683\_03 300001 410000  
WP AAI9683\_04 400001 510000  
WP AAI9683\_05 500001 610000  
WP AAI9683\_06 600001 710000  
WP AAI9683\_07 700001 810000  
WP AAI9683\_08 800001 910000  
WP AAI9683\_09 900001 1010000  
WP AAI9683\_10 1000001 1110000  
WP AAI9683\_11 1100001 1210000  
WP AAI9683\_12 1200001 1310000  
WP AAI9683\_13 1300001 1410000  
WP AAI9683\_14 1400001 1510000  
WP AAI9683\_15 1500001 1610000  
WP AAI9683\_16 1600001 1710000  
WP AAI9683\_17 1700001 1810000  
WP AAI9683\_18 1800001 1910000  
WP AAI9683\_19 1900001 2010000  
WP AAI9683\_20 2000001 2110000  
WP AAI9683\_21 2100001 2210000  
WP AAI9683\_22 2200001 2310000  
WP AAI9683\_23 2300001 2410000  
WP AAI9683\_24 2400001 2510000  
WP AAI9683\_25 2500001 2610000  
WP AAI9683\_26 2600001 2710000  
WP AAI9683\_27 2700001 2810000

Query Match  
Best Local Similarity 100.0%;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Score 19; DB 4; Length 110000;  
Pred. No. 38;

QY

1 GTGACGCCGGCGATGACC 19  
|||||

DB

27060 GTGACGCCGGCGATGACC 27042  
|||||

RESULT 4

AAI37969  
ID ACT137969 standard; DNA; 25 BP.  
XX AC ACT137969;  
DT 13-OCT-2003 (first entry)  
XX DE Human microarray DNA oligonucleotide SEQ ID NO 37960.  
XX EST; ss; probe; expressed sequence tag; microarray; gene expression;  
KW Genetic variation; biallelic marker; polymorphism; human;  
XX cross-species comparison.  
OS Homo sapiens.  
XX US2003104410-A1.  
XX PD 05-JUN-2003.  
XX PF 15-MAR-2002; 2002US-00098263.  
XX PR 16-MAR-2001; 2001US-0276759P.  
XX PA (AFFY-) AFFYMETRIX INC.  
XX PI Mittmann MP;  
XX DR WPI; 2003-567953/53.  
XX PT New array of nucleic acid probes, useful for in situ hybridization, in Southern, Northern or dot-blot hybridization to identify or detect the sequence or specific mutations of any gene.  
XX PS Claim 1; SEQ ID NO 37960; 9pp; English.  
XX CC The invention discloses a microarray comprising a plurality of nucleic acid probes including one of 2,018,500 fully defined sequences, or its perfect match, antisense match or antisense mismatch. The array is used also disclosed is a method of gene expression analysis. The array is used in monitoring gene expression levels by hybridisation to a DNA library, in analysis of genetic variation or in hybridisation of tag-labelled compounds. The nucleic acid probes are specifically designed for analysis of at least one target sequence. The method of analysis comprises hybridising at least one or more nucleic acids to at least two or more nucleic acid probes and detecting the hybridisation. The nucleic acid probes are attached to a solid support. The analysis comprises monitoring

CC gene expression levels, identifying biallelic markers or polymorphisms,  
 CC or family members of a gene and a cross-species comparison. Each of the  
 CC nucleic acids further comprises a tag sequence. The array of nucleic acid  
 CC probes is useful in situ hybridisation, in Southern, Northern or dot-  
 CC blot hybridisation to identify or detect the sequence or specific  
 CC mutations of any gene, in mapping the 5' termini of mRNA molecules by  
 CC primer extensions or in screening cDNA or genomic libraries or subclones  
 CC for additional subclones containing segments of DNA that have been  
 CC isolated and previously sequenced. The sequence presented is one of the  
 CC nucleic acid probes incorporated in the microarray. Note: The sequence  
 CC data for this patent can also be obtained in electronic format directly  
 CC from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)

XX Sequence 25 BP; 4 A; 8 C; 9 G; 4 T; 0 U; 0 Other;

Query Match 89.5%; Score 17; DB 8; Length 25;

Best Local Similarity 100.0%; Pred. No. 4.6e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCCGCATGA 17

Db 9 GTCGACGCCCGCATGA 25

#### RESULT 5

ABL72618/c  
 ID ABL72618 standard; cDNA; 270 BP.

XX ABL72618;

XX 14-MAY-2002 (first entry)

DE Corn tassell-derived polynucleotide (cdps) SEQ ID NO:1992.

XX Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPs;  
 KW inheritance; characteristic; growth; development; disease resistance;  
 KW environmental adaptability; quality; yield; molecular marker;  
 KW multigene trait; plant breeding; corn tassell; gene; ss.

XX Zea mays.

XX US2001051335-A1.

XX 13-DEC-2001.

XX 16-APR-1999; 99US-00294093.

XX 21-APR-1998; 99US-0082567P.

XX (LALG/) LALGUDI R V.

XX (ITOL/) ITO L Y.

XX (SHER/) SHERMAN B K.

XX Lalgudi RV, Ito LY, Sherman BK;

XX WPI; 2002-163647/21.

DR Novel purified corn tassell-derived polynucleotide useful for determining  
 PT altered gene expression, to recover regulatory elements and to follow  
 PT inheritance of desirable characteristics through hybrid breeding  
 PT programs.

XX Claim 1; SEQ ID NO 1992; 201pp; English.

XX The present sequence describes a purified corn tassell-derived  
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
 CC encode corn tassell-derived polypeptides (CDPs). The cdps sequences (1)  
 CC can be used for determining altered gene expression, to recover  
 CC regulatory elements and to follow inheritance of desirable  
 CC characteristics through hybrid breeding programs. (1) are also useful in  
 CC the evaluation, and alteration of desired characteristics associated with  
 CC growth and development, disease resistance, environmental adaptability,

CC quality and yield, and as molecular markers for studying inheritance of  
 CC multigene traits in a plant breeding program. (1) can be used to produce  
 CC a tassell-specific profile of gene transcription, a transcript image, to  
 CC clone regulatory elements for use in transformation vectors, to express a  
 CC polypeptide, to identify, isolate or extend identical or related corn  
 CC tassell nucleic acid sequences from DNA libraries, in nucleic acid  
 CC hybridisation or amplification technologies, as query sequences to  
 CC determine homology of known sequences, as probe for use in Southern or  
 CC Northern hybridisation, and to identify the presence of and/or to  
 CC determine the degree of similarity between two (or more) nucleic acid  
 CC sequences

XX Sequence 270 BP; 40 A; 97 C; 82 G; 50 T; 0 U; 1 Other;

Query Match 86.3%; Score 16.4; DB 6; Length 270;

Best Local Similarity 94.4%; Pred. No. 7.2e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCCGCATGAC 18

Db 87 GTCGACGCCCGCATGAC 70

#### RESULT 6

AAC39698/c

ID AAC39698 standard; DNA; 432 BP.

XX AAC39698;

XX 17-OCT-2000 (first entry)

DE Zea mays DNA fragment SEQ ID NO: 25575.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway; metabolic; pathway;

XX promoter; termination sequence; corn; ss.

XX Zea mays subsp. mays.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 03-MAR-1999; 99US-0123180P.

XX 23-MAR-1999; 99US-0123548P.

XX 25-MAR-1999; 99US-0125788P.

XX 01-APR-1999; 99US-0126264P.

XX 06-APR-1999; 99US-0127462P.

XX 08-APR-1999; 99US-0128234P.

XX 16-APR-1999; 99US-0128714P.

XX 19-APR-1999; 99US-0130077P.

XX 21-APR-1999; 99US-0130449P.

XX 23-APR-1999; 99US-0130510P.

XX 28-APR-1999; 99US-0130891P.

XX 30-APR-1999; 99US-0131449P.

XX 30-APR-1999; 99US-0132048P.

XX 04-MAY-1999; 99US-0132407P.

XX 05-MAY-1999; 99US-0132484P.

XX 06-MAY-1999; 99US-0132486P.

XX 07-MAY-1999; 99US-0132487P.

XX 11-MAY-1999; 99US-0132863P.

XX 14-MAY-1999; 99US-0134218P.

XX 14-MAY-1999; 99US-0134219P.

XX 14-MAY-1999; 99US-0134221P.

XX 18-MAY-1999; 99US-0134370P.

XX 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.  
 PR 20-MAY-1999; 99US-0135124P.  
 PR 21-MAY-1999; 99US-0135353P.  
 PR 24-MAY-1999; 99US-0135629P.  
 PR 25-MAY-1999; 99US-0136021P.  
 PR 27-MAY-1999; 99US-0136392P.  
 PR 28-MAY-1999; 99US-0136782P.  
 PR 01-JUN-1999; 99US-0137222P.  
 PR 03-JUN-1999; 99US-0137520P.  
 PR 04-JUN-1999; 99US-0137502P.  
 PR 07-JUN-1999; 99US-0137724P.  
 PR 08-JUN-1999; 99US-0138094P.  
 PR 10-JUN-1999; 99US-0138540P.  
 PR 14-JUN-1999; 99US-0138847P.  
 PR 16-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139492P.  
 PR 18-JUN-1999; 99US-0139492P.  
 PR 18-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 21-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140695P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 29-JUN-1999; 99US-0140991P.  
 PR 30-JUN-1999; 99US-0141267P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 21-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145192P.  
 PR 23-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 23-JUL-1999; 99US-0145224P.  
 PR 26-JUL-1999; 99US-0145276P.  
 PR 26-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 05-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 06-AUG-1999; 99US-0147493P.  
 PR 08-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149929P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 23-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 27-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154019P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 05-OCT-1999; 99US-0157117P.  
 PR 08-OCT-1999; 99US-0157753P.  
 PR 08-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.

```
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161993P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      86.3%; Score 16.4; DB 3; Length 432;
Best Local Similarity 94.4%; Pred. NO. 7+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
      |||||
Db      421 TCGACGCCGCCGATGACC 404

RESULT 7
AAC41047/c
ID AAC41047 standard; DNA; 445 BP.
XX
AC AAC41047;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 30455.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic; pathway;
KW promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0125864P.
PR 29-MAR-1999; 99US-0125785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 15-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
```

```
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 13-SEP-1999; 99US-0154018P.
PR 15-SEP-1999; 99US-0154039P.
PR 16-SEP-1999; 99US-0154779P.
PR 20-SEP-1999; 99US-0155139P.
PR 22-SEP-1999; 99US-0155486P.
PR 23-SEP-1999; 99US-0155659P.
PR 24-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159337P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161559P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      86.3%; Score 16.4; DB 3; Length 445;
Best Local Similarity 94.4%; Pred. No. 7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19
Db 436 TCGACGCCGCCGATGACC 419

RESULT 8
ADA70332/C
ID ADA70332 standard; DNA; 963 BP.
XX AC ADA70332;
XX DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 3655.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX gene; ds.
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX DR WPI; 2003-175290/17.
XX PT Identifying at least one gene involved in plant resistance or response to
XX PT pathogenic infection for conferring resistance or tolerance to a plant to
XX PT bacterial, fungal or viral infection by determining or detecting plant
XX PT gene expression.
XX PS Claim 6; SEQ ID NO 3655; 899pp; English.
XX CC The present invention relates to a method (M1) for identifying genes
XX CC involved in plant resistance or response to pathogenic infection. M1
XX CC comprises identifying a gene whose expression is significantly altered in
XX CC the incompatible interaction of plant gene expression relative to
XX CC expression of the gene in an uninfected plant, in a mutant plant that
XX CC does not express a gene associated with response to pathogenic infection,
XX CC or in a corresponding incompatible or compatible interaction. (M1) is
XX CC useful for conferring resistance to resistance or tolerance to a plant to
XX CC bacterial, fungal or viral infection. The present sequence was used to
XX CC illustrate the invention.
XX SQ Sequence 963 BP; 159 A; 334 C; 292 G; 178 T; 0 U; 0 Other;

Query Match      86.3%; Score 16.4; DB 7; Length 963;
Best Local Similarity 94.4%; Pred. No. 6.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 501 GTCGACGCCGCCGATGAC 484

RESULT 9
ADA48367/C
ID ADA48367 standard; DNA; 1077 BP.
XX AC ADA48367;
```

```

XX 20-NOV-2003 (first entry)
XX Rice gene conferring disease resistance in plants.
DE disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.
XX Oryza sativa.
XX WO2003000906-A2.
XX 03-JAN-2003.
XX 21-JUN-2002; 2002WO-IB002453.
XX 22-JUN-2001; 2001US-0300112P.
XX 26-SEP-2001; 2001US-0352277P.
XX 22-MAR-2002; 2002US-0366535P.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;
XX Katagiri F, Krepes J, Provart N, Ricke D, Zhu T;
XX WPI; 2003-184052/18.
XX P-PSDB; ADA48368.
XX New polynucleotide comprising a plant nucleotide sequence having an open
XX reading frame that encodes a polypeptide associated with disease
XX resistance, useful for conferring resistance or tolerance to a plant
XX pathogen.
XX Claim 1; SEQ ID NO 437; 299pp; English.
XX The invention relates to a novel isolated polynucleotide comprising a
XX plant nucleotide sequence having an open reading frame that encodes a
XX polypeptide associated with disease resistance or its fragment having
XX substantially the same activity as the full-length polypeptide. The
XX polynucleotide of the invention is useful for conferring resistance or
XX tolerance to a plant pathogen. The present sequence represents a gene
XX conferring disease resistance used in the invention.
XX Sequence 1077 BP; 189 A; 368 C; 363 G; 157 T; 0 U; 0 Other;
XX Query Match 86.3%; Score 16.4; DB 8; Length 1077;
XX Best Local Similarity 94.4%; Pred. NO. 6.6e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 GTCGACGCCGCCGATGAC 18
XX 171 GTCGACGTCGCCGATGAC 154
XX Db
XX RESULT 10
XX AAD29838/c
XX ID AAD29838 standard; DNA; 1560 BP.
XX AC AAD29838;
XX 17-May-2002 (first entry)
XX Rice NADPH dependent thioredoxin reductase DNA (NTR).
XX Transgenic plant; thioredoxin reductase; starch; protein; grain;
XX milling process; enzyme; rice; ds.
XX Oryza sativa.
XX Key Location/Qualifiers
XX CDS 406..1338
XX FT /*tag=a
XX FT /product="Rice NTR"
XX FT
XX

```

```

PN WO200198509-A2.
XX 27-DEC-2001.
XX 19-JUN-2001; 2001WO-EP006918.
XX 21-JUN-2000; 2000US-00598747.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX Lanahan MB, Desai NM, Gasdaska PY;
XX WPI; 2002-179557/23.
XX P-PSDB; AAE18733.
XX Transgenic plant coding for eukaryotic thioredoxin reductase at elevated
XX levels useful for separating the starch and protein components of grain
XX in a milling process.
XX Claim 12; Page 81; 86pp; English.
XX The present invention relates to a transgenic plant comprising
XX heterologous DNA coding for eukaryotic thioredoxin reductase integrated
XX into its nuclear or plastid genome and use of thioredoxin reductase for
XX separating the starch and protein components of grain in a milling
XX process. Transgenic plant is used for separating the starch and protein
XX components of grain in a milling process. Transgenic plant may be used to
XX produce thioredoxin reductase at elevated levels. Delivery of thioredoxin
XX reductase eliminates the need to develop exogenous sources for addition
XX during processing. Secondly, physical disruption of seed integrity is not
XX necessary to bring the enzyme in contact with the storage or matrix
XX proteins of the seed prior to processing or as an extra processing step.
XX The present sequence is rice NADPH dependent thioredoxin reductase DNA
XX (NTR)
XX SQ Sequence 1560 BP; 302 A; 435 C; 439 G; 372 T; 0 U; 12 Other;
XX Query Match 86.3%; Score 16.4; DB 6; Length 1560;
XX Best Local Similarity 94.4%; Pred. NO. 6.4e+02;
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 GTCGACGCCGCCGATGAC 18
XX 906 GTCGACGCCGCCGATGAC 889
XX Db
XX RESULT 11
XX ACA27231/c
XX ID ACA27231 standard; DNA; 1948 BP.
XX AC ACA27231;
XX 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #8888.
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX Bordetella pertussis.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US0009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362899P.
XX

```

PA (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR P-PSDB; ABU23361.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 14; SEQ ID NO 15101; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1948 BP; 289 A; 684 C; 682 G; 293 T; 0 U; 0 Other;  
 Query Match 86.3%; Score 16.4; DB 7; Length 1948;  
 Best Local Similarity 94.4%; Pred. No. 6.3e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCGACGCCGCCGATGAC 18  
 Db 96 GTCCGCCGCCGATGAC 79  
 RESULT 12  
 ABZ16341  
 ID ABZ16341 standard; DNA; 2000 BP.  
 XX  
 AC ABZ16341;  
 XX  
 DT 21-JAN-2003 (first entry)  
 XX  
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4146.  
 XX  
 KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200216655-A2.  
 XX

PD 28-FEB-2002.  
 XX  
 PF 24-AUG-2001; 2001WO-US026685.  
 XX  
 PR 24-AUG-2000; 2000US-0227866P.  
 PR 26-JAN-2001; 2001US-0264647P.  
 PR 22-JUN-2001; 2001US-0300111P.  
 XX  
 PA (SCRI ) SCRIPPS RES INST.  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Harper JF, Kreps J, Wang X, Zhu T;  
 PI WPI; 2002-304127/34.  
 DR  
 XX Identifying a stress condition to which a plant cell has been exposed and  
 PT producing plants with increased tolerance to these abiotic stresses.  
 PT  
 PS Claim 14; SEQ ID NO 4146; 577pp + Sequence Listing; English.  
 XX  
 CC The invention relates to identifying a stress condition to which a plant  
 CC cell has been exposed, comprising: (a) contacting nucleic acid  
 CC representative of expressed polynucleotides in the plant cell with an  
 CC array or probes representative of the plant cell genome; and (b)  
 CC detecting a profile of expressed polynucleotides in the plant cell  
 CC characteristic of a stress response. The method is useful in the  
 CC production of transgenic plants, cells and seeds and in producing plants  
 CC with increased tolerance to abiotic stress. The present sequence is that  
 CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used  
 CC in methods of the invention. Note: The sequence data for this patent is  
 CC not represented in the printed specification but is based on sequence  
 CC information supplied to Derwent by the European Patent Office  
 XX  
 SQ Sequence 2000 BP; 683 A; 368 C; 282 G; 667 T; 0 U; 0 Other;  
 Query Match 86.3%; Score 16.4; DB 6; Length 2000;  
 Best Local Similarity 94.4%; Pred. No. 6.3e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCGACGCCGCCGATGAC 18  
 Db 353 GTCCGCCGCCGATGAC 370  
 RESULT 13  
 AAA81465  
 ID AAA81465 standard; DNA; 47475 BP.  
 XX  
 AC AAA81465;  
 XX  
 DT 04-DEC-2000 (first entry)  
 XX  
 DE N. meningitidis partial DNA sequence gnm\_13 SEQ ID NO:13.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200022430-A2.  
 XX  
 PD 20-APR-2000.  
 XX  
 PF 08-OCT-1999; 99WO-US023573.  
 XX  
 PR 09-OCT-1998; 98US-0103794P.  
 PR 30-APR-1999; 99US-0132068P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;  
 PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;



PI Rappuoli R, Pizza M;  
XX WPI; 2000-318079/27.  
DR  
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be used  
PT in the diagnosis and treatment of *N. meningitidis* infection and other  
PT *Neisseria* infections, for example, *N. gonorrhoea*.  
XX  
XX  
PS Claim 7; Page 383-397; 1760pp; English.  
XX  
XX The present invention describes methods of obtaining immunogenic proteins  
CC from *Neisseria* genomic sequences. AA81453 to AA82414 represent  
CC specifically claimed *Neisseria meningitidis* genomic DNA sequences;  
CC AA81260 to AA81303 and AA825620 to AA825663 represent *Neisseria* DNA  
CC sequences and their corresponding proteins; AA81254 to AA81259 and  
CC AA81304 to AA81321 represent PCR primers used in the isolation of  
CC *Neisseria meningitidis* DNA sequences; and AA81322 to AA81452 represent  
CC *Neisseria meningitidis* MenB polynucleotide ORF sequences, which are all  
CC used in the exemplification of the present invention. The nucleic acid  
CC sequences, protein sequences, and antibodies against them, can be used as a  
CC the manufacture of a composition. The composition can be used as a  
CC medicament (or in the manufacture of a medicament) for treating,  
CC preventing or diagnosing infection due to *Neisseria* bacteria. For  
CC example, some of the identified proteins could be components of vaccines  
CC against *Meningococcus* B; against all serotypes; and/or against all  
CC pathogenic *Neisseria*. Identification of sequences from the bacterium  
CC will also facilitate production of biological probes, particularly  
CC organism-specific probes. Attempts to make efficacious *Meningococcus* B  
CC vaccines have failed mainly due to antigen tolerance. Multivalent  
CC vaccines have also been tried but none have successfully overcome  
CC antigenic variability. The provision of further, complete sequences may  
CC provide an opportunity to identify secreted or surface exposed proteins  
CC that may be presumed targets for the immune system and which are not  
CC antigenically variable or at least more conserved than other more  
CC variable regions  
XX  
SQ Sequence 47475 BP; 11600 A; 13758 C; 12290 G; 9826 T; 0 U; 1 Other;  
Query Match 86.3%; Score 16.4; DB 3; Length 47475;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCGACGCCCGCGATGAC 19  
DB 13955 TCGACGCCCGCGAGACC 13972  
RESULT 14  
AAA81489\_5  
Continuation (6 of 9) of AAA81489 from base 500001 (*N. meningitidis* partial DNA sequence  
WP Sequence split into 9 fragments LOCUS AAA81489 Accession AAA81489  
WP Fragment Name Begin End  
WP AAA81489\_0 1 110000  
WP AAA81489\_1 100001 210000  
WP AAA81489\_2 200001 310000  
WP AAA81489\_3 300001 410000  
WP AAA81489\_4 400001 510000  
WP AAA81489\_5 500001 610000  
WP AAA81489\_6 600001 710000  
WP AAA81489\_7 700001 810000  
WP AAA81489\_8 800001 837096  
Query Match 86.3%; Score 16.4; DB 3; Length 110000;  
Best Local Similarity 94.4%; Pred. No. 4.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCGACGCCCGCGATGAC 19  
DB 23196 TCGACGCCCGCGAGACC 23213  
RESULT 15  
AA199682\_02

Continuation (3 of 45) of AA199682 from base 200001 (*Mycobacterium tuberculosis* strain H  
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682  
WP Fragment Name Begin End  
WP AA199682\_00 1 110000  
WP AA199682\_01 100001 210000  
WP AA199682\_02 200001 310000  
WP AA199682\_03 300001 410000  
WP AA199682\_04 400001 510000  
WP AA199682\_05 500001 610000  
WP AA199682\_06 600001 710000  
WP AA199682\_07 700001 810000  
WP AA199682\_08 800001 910000  
WP AA199682\_09 900001 1010000  
WP AA199682\_10 1000001 1110000  
WP AA199682\_11 1100001 1210000  
WP AA199682\_12 1200001 1310000  
WP AA199682\_13 1300001 1410000  
WP AA199682\_14 1400001 1510000  
WP AA199682\_15 1500001 1610000  
WP AA199682\_16 1600001 1710000  
WP AA199682\_17 1700001 1810000  
WP AA199682\_18 1800001 1910000  
WP AA199682\_19 1900001 2010000  
WP AA199682\_20 2000001 2110000  
WP AA199682\_21 2100001 2210000  
WP AA199682\_22 2200001 2310000  
WP AA199682\_23 2300001 2400001  
WP AA199682\_24 2400001 2510000  
WP AA199682\_25 2500001 2610000  
WP AA199682\_26 2600001 2710000  
WP AA199682\_27 2700001 2810000  
WP AA199682\_28 2800001 2910000  
WP AA199682\_29 2900001 3010000  
WP AA199682\_30 3000001 3110000  
WP AA199682\_31 3100001 3210000  
WP AA199682\_32 3200001 3310000  
WP AA199682\_33 3300001 3410000  
WP AA199682\_34 3400001 3510000  
WP AA199682\_35 3500001 3610000  
WP AA199682\_36 3600001 3710000  
WP AA199682\_37 3700001 3810000  
WP AA199682\_38 3800001 3910000  
WP AA199682\_39 3900001 4010000  
WP AA199682\_40 4000001 4110000  
WP AA199682\_41 4100001 4210000  
WP AA199682\_42 4200001 4310000  
WP AA199682\_43 4300001 4410000  
WP AA199682\_44 4400001 4411529  
Query Match 86.3%; Score 16.4; DB 4; Length 110000;  
Best Local Similarity 94.4%; Pred. No. 4.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTCGACGCCCGCGATGAC 18  
DB 77268 GACGACGCCCGCGATGAC 77285  
RESULT 16  
AA199683\_02  
Continuation (3 of 44) of AA199683 from base 200001 (*Mycobacterium tuberculosis* strain H  
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683  
WP Fragment Name Begin End  
WP AA199683\_00 1 110000  
WP AA199683\_01 100001 210000  
WP AA199683\_02 200001 310000  
WP AA199683\_03 300001 410000  
WP AA199683\_04 400001 510000  
WP AA199683\_05 500001 610000  
WP AA199683\_06 600001 710000  
WP AA199683\_07 700001 810000  
WP AA199683\_08 800001 910000  
WP AA199683\_09 900001 1010000

WP AAI99683\_10 1110000  
 WP AAI99683\_11 1210000  
 WP AAI99683\_12 1310000  
 WP AAI99683\_13 1410000  
 WP AAI99683\_14 1510000  
 WP AAI99683\_15 1610000  
 WP AAI99683\_16 1710000  
 WP AAI99683\_17 1810000  
 WP AAI99683\_18 1910000  
 WP AAI99683\_19 2010000  
 WP AAI99683\_20 2110000  
 WP AAI99683\_21 2210000  
 WP AAI99683\_22 2310000  
 WP AAI99683\_23 2410000  
 WP AAI99683\_24 2510000  
 WP AAI99683\_25 2610000  
 WP AAI99683\_26 2710000  
 WP AAI99683\_27 2810000  
 WP AAI99683\_28 2910000  
 WP AAI99683\_29 3010000  
 WP AAI99683\_30 3110000  
 WP AAI99683\_31 3210000  
 WP AAI99683\_32 3310000  
 WP AAI99683\_33 3410000  
 WP AAI99683\_34 3510000  
 WP AAI99683\_35 3610000  
 WP AAI99683\_36 3710000  
 WP AAI99683\_37 3810000  
 WP AAI99683\_38 3910000  
 WP AAI99683\_39 4010000  
 WP AAI99683\_40 4110000  
 WP AAI99683\_41 4210000  
 WP AAI99683\_42 4310000  
 WP AAI99683\_43 4403765

Query Match 86.3%; Score 16.4; DB 4; Length 110000;  
 Best Local Similarity 94.4%; Pred. No. 4.9e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTCGACGCCGCGATGAC 18  
 Db 77380 GACGACGCCGCGATGAC 77397

RESULT 17  
 ID AAF21612  
 AC AAF21612;  
 DT 13-MAR-2001 (first entry)

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.

XX Neisseria meningitidis.

OS WO200066791-A1.

FN 09-NOV-2000.

PD 08-MAR-2000; 2000WO-US005928.

PF 30-APR-1999; 99US-0132068P.

PR 08-OCT-1999; 99WO-US023573.

PR 28-FEB-2000; 2000GB-00004695.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC,  
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,  
 XX Rappuoli R, Frazer CM, Grandi G;  
 DR WPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent Neisserial infections.

PS Claim 7; Appendix A; 692pp; English.

XX The present invention describes the full length genome of Neisseria  
 CC meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613  
 CC represent fragments of the NMB genomic sequence, as the sequence was too  
 CC long to go in a record on its own it was split into 8 sequences which  
 CC overlap each other at the beginning and end of each sequence by 49980 bp  
 CC (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of  
 CC AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of  
 CC AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins  
 CC primers which are used in the exemplification of the present invention.  
 CC The NMB genome and fragments from it have antibacterial activity, and can  
 CC be used in vaccines and gene therapy. Neisseria nucleic acids, proteins  
 CC and/or antibodies which binds to the proteins can be used in compositions  
 CC for treating or preventing infection due to Neisserial bacteria or as a  
 CC diagnostic reagent for detecting the presence of Neisserial bacteria or  
 CC of antibodies raised to Neisserial bacteria. Computers, computer memory,  
 CC computer storage medium or computer databases can be used in a search to  
 CC identify open reading frames (ORFs) or coding sequences within the NMB  
 CC genome. The DNA sequences provide further opportunities to find antigenic  
 CC or immunogenic proteins which are more effective in vaccines than the  
 CC outer membrane proteins currently used

XX Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;

Query Match 86.3%; Score 16.4; DB 3; Length 349980;

Best Local Similarity 94.4%; Pred. No. 4.5e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCGACGCCGCGATGACC 19

Db 158425 TCGACGCCGCGAGACC 158442

RESULT 18

ABQ81846

ID ASQ81846 standard; DNA; 349980 BP.

XX AC ABQ81846;

XX DT 19-NOV-2002 (first entry)

XX Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1102.

XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

KW antiarrhythmic; antibacterial; inhibitor of salmonella; detection;

KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;

KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX Bifidobacterium longum.

OS Synthetic.

XX EP1227152-A1.

PN 31-JUL-2002.

XX 30-JAN-2001; 2001EP-00102050.

XX 30-JAN-2001; 2001EP-00102050.

XX (NEST) SOC PROD NESTLE SA.

XX WPI; 2002-668397/72.

XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
PT a probe or primer for detecting and/or identifying Bifidobacterium longum  
PT in a biological sample.  
XX  
XX Disclosure; SEQ ID NO 1102; 80pp; English.  
XX  
XX The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in AB081842 and AB081843, or a sequence exhibiting at  
CC least 90% identity or which hybridizes with the sequences given in  
CC AB081842 and AB081843. Also described is a polynucleotide (II) encoding a  
CC fusion protein, comprising a sequence selected from 1097 sequences given  
CC in ABP65258 to ABP65354 ligated in frame to a polynucleotide encoding a  
CC heterologous polypeptide. (I) has antidiarrheic and antibacterial  
CC activities, and can be used as an inhibitor of Salmonella. (I) (which is  
CC a probe) is useful for the detection and/or identification of  
CC Bifidobacterium longum in a biological sample. A carrier containing the  
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be  
CC used for preventing and/or treating diarrhoea brought about by pathogenic  
CC bacteria and/or rotavirus. The carrier is a food composition selected  
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented  
CC products, ice-creams, fermented cereal based products, milk based  
CC powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent  
CC Bifidobacterium related nucleotide sequences given in the Sequence  
CC Listing from the present invention but not mentioned further within the  
CC specification. N.B. The sequence data for this patent is not represented  
CC in the printed specification but is based on sequence information  
CC supplied by the European Patent Office  
XX  
XX SQ Sequence 349980 BP; 69200 A; 103414 C; 105147 G; 72219 T; 0 U; 0 Other;  
Query Match 86.3%; Score 16.4; DB 6; Length 349980;  
Best Local Similarity 94.4%; Pred. No. 4.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCCGATG 18  
DB 269605 GTCGACGCCGCCGATG 269622  
RESULT 19  
ID ADC23768/c  
XX ADC23768 standard; DNA; 942 BP.  
XX  
XX AC ADC23768;  
XX  
XX DT 18-DEC-2003 (first entry)  
XX  
XX DE DNA sequence (SeqID 35) encoding a nitrilase enzyme.  
XX  
XX KW gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
XX KW enantiomer; chiral medicine.  
XX  
XX OS Unidentified.  
XX  
XX OS WO2003000840-A2.  
XX  
XX PN 03-JAN-2003.  
XX  
XX PF 15-MAY-2002; 2002WO-US015983.  
XX  
XX PR 21-JUN-2001; 2001US-0300189P.  
XX PR 30-JUL-2001; 2001US-030906P.  
XX PR 22-JAN-2002; 2002US-0351336P.  
XX  
XX XX (DIVE-) DIVERSA CORP.  
XX PA (MADD-) MADDEN D.  
XX  
XX DR

PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX WPI; 2003-201417/19.  
DR P-P8DB; ADC23769.  
XX  
XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
XX Claim 1; SEQ ID NO 35; 560pp; English.  
XX  
XX This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polynucleotide is a DNA  
CC sequence that encodes a nitrilase enzyme of the invention.  
XX  
XX SQ Sequence 942 BP; 220 A; 221 C; 275 G; 226 T; 0 U; 0 Other;  
Query Match 84.2%; Score 16; DB 9; Length 942;  
Best Local Similarity 100.0%; Pred. No. 9.9e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCCGATG 16  
DB 384 GTCGACGCCGCCGATG 369  
RESULT 20  
ID AAA53972  
XX AAA53972 standard; DNA; 1821 BP.  
XX  
XX AC AAA53972;  
XX  
XX DT 08-FEB-2001 (first entry)  
XX  
XX DE M. tuberculosis Mtb-67.2 antigen.  
XX  
XX KW Mycobacterium tuberculosis; antigen; Mtb-91; Mtb-67.2; APC;  
XX KW antigen presenting cell; serodiagnosis; detection;  
XX KW human immunodeficiency virus; HIV; acquired immune deficiency syndrome;  
XX KW AIDS; ds.  
XX  
XX OS Mycobacterium tuberculosis.  
XX  
XX PH Key Location/Qualifiers  
XX FT CDS 1..1821  
XX FT /\*tag= a  
XX FT /product= "Mtb-67.2 antigen"  
XX  
XX PN WO2000055194-A2.  
XX  
XX PN 21-SEP-2000.  
XX  
XX PF 17-MAR-2000; 2000WO-US007196.  
XX  
XX PR 18-MAR-1999; 99US-00272975.  
XX  
XX PR (CORI-) CORIXA CORP.  
XX  
XX PI Hendrickson RC, Lodes MJ, Houghton RL;  
XX WPI; 2000-638180/61.  
XX  
XX DR

DR P-PSDB; AAB00118.  
XX Novel Mycobacterium tuberculosis polypeptide comprising an immunogenic  
PT portion of M. tuberculosis antigens Mtb-81 and Mtb-67.2, useful for  
PT diagnosis, treatment and monitoring therapy of tuberculosis.  
XX  
PS Claim 83; Fig 4; 91pp; English.  
XX  
XX Polypeptides comprising an immunogenic portion of Mycobacterium  
CC tuberculosis antigens Mtb-81 or Mtb-67.2 or variants that differ by  
CC substitutions, additions, insertions and/or deletions but which still  
CC react with antigen specific antisera or T-cells are described. Also  
CC described are polynucleotides encoding the polypeptides. The  
CC polypeptides, expression vectors expressing the polypeptides or  
CC comprising an antisense polynucleotide, or an antigen presenting cell  
CC comprising a sequence encoding the polypeptides are useful for  
CC determining the presence or absence of M. tuberculosis in whole blood,  
CC serum, sputum, plasma, saliva, cerebrospinal fluid or urine in a patient  
CC infected with human immunodeficiency virus (HIV)  
XX  
SQ Sequence 1821 BP; 350 A; 566 C; 601 G; 304 T; 0 U; 0 Other;  
Query Match 84.2%; Score 16; DB 3; Length 1821;  
Best Local Similarity 100.0%; Pred. No. 9.5e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCCGATG 16  
Db 1666 GTCGACGCCGCCGATG 1681  
RESULT 21  
AAI75554  
ID AAI75554 standard; DNA; 51 BP.  
XX  
AC AAI75554;  
XX  
DT 09-NOV-2001 (first entry)  
XX  
DE Human silent SNP containing nucleic acid SEQ:2495.  
XX  
KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;  
KW protein therapy; vaccine; probe; diagnostic assay; detection;  
KW quantitation; restorative therapy; polymorphic; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200140521-A2.  
XX  
PD 07-JUN-2001.  
XX  
PF 30-NOV-2000; 2000WO-US032758.  
XX  
PR 30-NOV-1999; 99US-0168138P.  
PR 29-NOV-2000; 2000US-00726173.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI; 2001-356160/37.  
XX  
XX Polymorphic nucleic acid sequences, useful in genetic testing and  
PT therapy.  
PS Claim 1; Page 815; 2653pp; English.  
XX  
XX AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide  
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).  
CC AAM53114 to AAM53329 represent peptides related to human polymorphic  
CC polynucleotide sequences. The sequences can be used in gene and protein  
CC therapy, and in vaccine production. (I) and the polypeptides encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
XX associated with inappropriate expression of polymorphic polypeptides. For  
XX example, (I) may be used to treat disorders by rectifying mutations or  
XX deletions in a patient's genome that affect the activity of polypeptides  
XX by expressing inactive proteins or to supplement the patients own  
XX production of polypeptide. Additionally, (I) and its complementary  
XX sequences may also be used as DNA probes in diagnostic assays to detect

CC associated with inappropriate expression of polymorphic polypeptides. For  
CC example, (I) may be used to treat disorders by rectifying mutations or  
CC deletions in a patient's genome that affect the activity of polypeptides  
CC by expressing inactive proteins or to supplement the patients own  
CC production of polypeptide. Additionally, (I) and its complementary  
CC sequences may also be used as DNA probes in diagnostic assays to detect  
CC and quantitate the presence of similar nucleic acids in samples, and  
CC therefore which patients may be in need of restorative therapy. The  
CC polypeptides encoded by (I) may be used as antigens in the production of  
CC antibodies specific for polymorphic polypeptides. The antibodies may also  
CC be used to down regulate expression and activity. The antibodies may also  
CC be used as diagnostic agents for detecting the presence of polymorphic  
CC polypeptides in samples  
XX  
SQ Sequence 51 BP; 7 A; 20 C; 17 G; 7 T; 0 U; 0 Other;  
Query Match 83.2%; Score 15.8; DB 4; Length 51;  
Best Local Similarity 89.5%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCCGATGACC 19  
Db 5 GCCGACGCCGCCGATGCC 23  
RESULT 22  
AAI75555  
ID AAI75555 standard; DNA; 51 BP.  
XX  
AC AAI75555;  
XX  
DT 09-NOV-2001 (first entry)  
XX  
DE Human silent SNP containing nucleic acid SEQ:2496.  
XX  
KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;  
KW protein therapy; vaccine; probe; diagnostic assay; detection;  
KW quantitation; restorative therapy; polymorphic; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200140521-A2.  
XX  
PD 07-JUN-2001.  
XX  
PF 30-NOV-2000; 2000WO-US032758.  
XX  
PR 30-NOV-1999; 99US-0168138P.  
PR 29-NOV-2000; 2000US-00726173.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI; 2001-356160/37.  
XX  
XX Polymorphic nucleic acid sequences, useful in genetic testing and  
PT therapy.  
PS Claim 1; Page 815; 2653pp; English.  
XX  
XX AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide  
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).  
CC AAM53114 to AAM53329 represent peptides related to human polymorphic  
CC polynucleotide sequences. The sequences can be used in gene and protein  
CC therapy, and in vaccine production. (I) and the polypeptides encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
XX associated with inappropriate expression of polymorphic polypeptides. For  
XX example, (I) may be used to treat disorders by rectifying mutations or  
XX deletions in a patient's genome that affect the activity of polypeptides  
XX by expressing inactive proteins or to supplement the patients own  
XX production of polypeptide. Additionally, (I) and its complementary  
XX sequences may also be used as DNA probes in diagnostic assays to detect

CC and quantitate the presence of similar nucleic acids in samples, and  
CC therefore which patients may be in need of restorative therapy. The  
CC polypeptides encoded by (I) may be used as antigens in the production of  
CC antibodies specific for polymorphic polypeptides. The antibodies may also  
CC be used to down regulate expression and activity. The antibodies may also  
CC be used as diagnostic agents for detecting the presence of polymorphic  
CC polypeptides in samples

XX SQ Sequence 51 BP; 6 A; 21 C; 17 G; 7 T; 0 U; 0 Other;  
Query Match 83.2%; Score 15.8; DB 4; Length 51;  
Best Local Similarity 89.5%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCGACGCGCGCGATGACC 19  
DB 5 GCGGACGCGCGCGATGCC 23

RESULT 23  
AAI79702  
ID AAI79702 standard; DNA; 51 BP.  
XX AC AAI79702;  
XX DT 09-NOV-2001 (first entry)  
XX DE Human conservative amino acid changing SNP nucleic acid SEQ:6643.  
XX KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;  
XX KW protein therapy; vaccine; probe; diagnostic assay; detection;  
XX KW quantitation; restorative therapy; polymorphic; ds.  
XX OS Homo sapiens.  
XX PN WO200140521-A2.  
XX PD 07-JUN-2001.  
XX PF 30-NOV-2000; 2000WO-US032758.  
XX PR 30-NOV-1999; 99US-0168138P.  
XX PR 29-NOV-2000; 2000US-00726173.  
XX PA (CURA-) CURAGEN CORP.  
XX PI Shimkets RA, Leach M;  
XX PT WPI; 2001-356160/37.  
XX PT Polymorphic nucleic acid sequences, useful in genetic testing and  
XX PT therapy.  
XX PS Claim 1; Page 2539; 2653pp; English.

CC AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide  
CC sequences (I), which contain single nucleotide polymorphisms (SNPs).  
CC AAI73114 to AAI73329 represent peptides related to human polymorphic  
CC polynucleotide sequences. The sequences can be used in gene and protein  
CC therapy, and in vaccine production. (I) and the polypeptides encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate expression of polymorphic polypeptides. For  
CC example, (I) may be used to treat disorders by rectifying mutations or  
CC deletions in a patient's genome that affect the activity of polypeptides  
CC by expressing inactive proteins or to supplement the patients own  
CC production of polypeptide. Additionally, (I) and its complementary  
CC sequences may also be used as DNA probes in diagnostic assays to detect  
CC and quantitate the presence of similar nucleic acids in samples, and  
CC therefore which patients may be in need of restorative therapy. The  
CC polypeptides encoded by (I) may be used as antigens in the production of  
CC antibodies specific for polymorphic polypeptides. The antibodies may also  
CC be used to down regulate expression and activity. The antibodies may also  
CC be used as diagnostic agents for detecting the presence of polymorphic

CC polypeptides in samples

XX SQ Sequence 51 BP; 6 A; 22 C; 14 G; 9 T; 0 U; 0 Other;  
Query Match 83.2%; Score 15.8; DB 4; Length 51;  
Best Local Similarity 89.5%; Pred. No. 1.4e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCGACGCGCGCGATGACC 19  
DB 19 GTCGACTCCCGCGATGACC 37

RESULT 24  
ACA42539/c  
ID ACA42539 standard; DNA; 477 BP.  
XX AC ACA42539;  
XX DT 19-JUN-2003 (first entry)  
XX DE Prokaryotic essential gene #24196.  
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX KW drug design; gene.  
XX OS Pseudomonas aeruginosa.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX PI WPI; 2003-029926/02.  
XX DR P-PSDB; ABU38669.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 14; SEQ ID NO 30409; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway;  
CC (8) required for proliferation, or that inhibits cellular proliferation;  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 477 BP; 81 A; 170 C; 150 G; 76 T; 0 U; 0 Other;  
 Query Match 83.2%; Score 15.8; DB 7; Length 477;  
 Best Local Similarity 89.5%; Pred. No. 1.3e-03;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGACC 19  
 Db 114 GTCGACGCGCGCGATGACC 96

RESULT 25  
 AAC75478  
 ID AAC75478 standard; cDNA; 598 BP.

AC AAC75478;  
 XX  
 XX 08-FEB-2001 (first entry)

XX Human ORFX ORF1033 polynucleotide sequence SEQ ID NO:2065.  
 DE  
 XX

KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnery; antipariatic; antiparkinsonian; nontropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive; ss.

OS Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US008621.

XX 31-MAR-1999; 99US-0127607P.

XX 02-APR-1999; 99US-0127636P.

XX 05-APR-1999; 99US-0127728P.

XX 30-MAR-2000; 2000US-00540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX P-PSDB; AAB41269.

XX Novel nucleic acids and peptides derived from open reading frame X,  
 XX useful for treating e.g. cancers, proliferative disorders,  
 XX neurodegenerative disorders and cardiovascular disease.

PS Claim 5; Page 1559-1560; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
 CC antipariatic; antiparkinsonian; nontropic; neuroprotective; osteopathic;  
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;  
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;  
 CC dermatological; immunosuppressive; antiinflammatory; antibacterial;  
 CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The  
 CC sequences can be used for determining the presence of or predisposition  
 CC to, or preventing or treating pathological conditions associated with an  
 CC ORFX-associated disorder. The nucleic acids can be used to express ORFX  
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be  
 CC used to treat cancers, proliferative disorders, neurodegenerative  
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,  
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester  
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency  
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune  
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and  
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to  
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive

XX Sequence 598 BP; 99 A; 198 C; 172 G; 129 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 3; Length 598;  
 Best Local Similarity 89.5%; Pred. No. 1.2e-03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGACC 19

Db 25 GTCGACGCGCGCGATGACC 43

RESULT 26

ABN21241

ID ABN21241 standard; cDNA; 598 BP.

AC ABN21241;

XX 24-JUN-2002 (first entry)

XX Human ORFX polynucleotide sequence SEQ ID NO:10959.

XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis; gene; ss.

OS Homo sapiens.

XX WO200192523-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US010836.

XX 30-MAY-2000; 2000US-0206132P.

XX 29-AUG-2000; 2000US-0228716P.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach MD;

XX WPI; 2002-106308/14.

XX P-PSDB; ABP05489.

XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,





PS Claim 1; SEQ ID NO 844; 617pp; English.

CC The invention relates to a novel isolated nucleic acid derived from

CC Nicotiana benthamiana, Oryza sativa (rice), Saccharomyces cerevisiae

CC (yeast), Trichoderma harzianum (Hypocrea lixii) and Papaver rhoeas

CC (poppy), and a sequence that hybridises to them under conditions of low

CC stringency, where expression of the nucleic acid in a plant results in a

CC disease resistance phenotype. The polynucleotides of the invention

CC demonstrate anti-fungal activity and may be useful in conferring disease

CC resistance in a plant against phytopathogen such as Aspergillus flavus,

CC Gibberella fujikuroi and Gibberella zeae. Furthermore, the

CC polynucleotides may be useful to retrieve unknown sequences and in gene

CC shuffling or sexual PCR procedures. The current sequence is that of the

CC DNA of the invention which is homologous to that of the phytopathogen

CC resistance-related contig cDNAs.

XX SQ Sequence 620 BP; 101 A; 227 C; 192 G; 100 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 9; Length 620;

Best Local Similarity 89.5%; Pred. NO. 1.2e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGCGCGCGCGATGACC 19

Db 331 GACGCGCGCGATGACC 313

RESULT 29

ACA42466/c

ID ACA42466 standard; DNA; 738 BP.

XX AC ACA42466;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #24123.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; gene.

XX Pseudomonas aeruginosa.

OS WO20027183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU38596.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 30336; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of

CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological

CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of

CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required

CC for cellular proliferation to isolate candidate molecules for rational

CC drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

CC prokaryotic essential genes. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 738 BP; 131 A; 224 C; 245 G; 138 T; 0 U; 0 Other;

Query Match 83.2%; Score 15.8; DB 7; Length 738;

Best Local Similarity 89.5%; Pred. NO. 1.2e+03;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTGCGCGCGCGCGATGACC 19

Db 186 GTGCGCGCGCGCGATGACC 168

RESULT 30

ADC98771/c

ID ADC98771 standard; DNA; 738 BP.

XX AC ADC98771;

XX 01-JAN-2004 (first entry)

DE Pseudomonas aeruginosa pyrH DNA - SEQ ID 64.

XX adenine phosphoribosyltransferase; apt; uridylylate kinase; pyrH;

KW guanylate kinase; gmk; antibacterial; vaccine; bronchial infection;

KW sinusitis; meningitis; food preservative; disinfectant; ds; gene.

XX Pseudomonas aeruginosa.

OS WO2003035858-A2.

PN 01-MAY-2003.

XX 25-OCT-2002; 2002WO-CA001613.

XX 25-OCT-2001; 2001US-0337625P.

PR 26-OCT-2001; 2001US-0340534P.

PR 18-DEC-2001; 2001US-0341639P.

PR 18-DEC-2001; 2001US-0341825P.

PR 18-DEC-2001; 2001US-0342004P.

PR 20-DEC-2001; 2001US-0342559P.

XX (AFFI-) AFFINIUM PHARM INC.

XX Edwards A, Dharamsi A, Vedadi M, Alam MZ, Awrey D, Beattie B;

PI Canadian V, Domagala M, Houston S, Li Q, Mansoury K, Necakov S;

PI Ng I, Pinder B, Sheldrick B, Wrezel O;



DR WPI; 2003-421420/39.  
XX P-PSDB; ADC98772.  
PT Novel recombinant polypeptide useful for designing a modulator for the  
PT prevention or treatment of Streptococcus aureus related disease or  
PT disorder.  
XX  
PS Claim 136; SEQ ID NO 64; 252pp; English.  
XX  
CC The invention relates to a novel composition comprising an isolated,  
CC recombinant polypeptide. The polypeptide may comprise Staphylococcus  
CC aureus or Streptococcus pneumoniae adenine phosphoribosyltransferase  
CC (apt), uridylylate kinase (pyrH) isolated from Staphylococcus aureus,  
CC Streptococcus pneumoniae or Pseudomonas aeruginosa or S. pneumoniae  
CC guanylate kinase (gmk). The polypeptides of the invention demonstrate  
CC antibacterial activity whilst the composition may be useful for designing  
CC a modulator facilitating the prevention or treatment of an S. aureus or  
CC S. pneumoniae related disease or disorder, such as bronchial infection,  
CC sinusitis and meningitis. Furthermore, the composition may be utilised to  
CC generate a vaccine for protection against invasion of bacteria, as well  
CC as to develop food preservatives and surface disinfectants. The current  
CC sequence is that of the P. aeruginosa pyrH DNA of the invention.  
XX  
SQ Sequence 738 BP; 131 A; 225 C; 245 G; 137 T; 0 U; 0 Other;  
Query Match 83.2%; Score 15.8; DB 9; Length 738;  
Best Local Similarity 89.5%; Pred. No. 1.2e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCGATGACC 19  
Db 186 GTTGGCGCGCGATGACC 168

Search completed: June 20, 2004, 10:17:49  
Job time : 122.3 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 887.873 Seconds  
(without alignments)  
639.034 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19  
Sequence: 1 gtcgacgcgcgcgatgacc 19

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27533289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database :

- EST:\*
- 1: em\_estba:\*
  - 2: em\_esthum:\*
  - 3: em\_estin:\*
  - 4: em\_estma:\*
  - 5: em\_estov:\*
  - 6: em\_estpl:\*
  - 7: em\_estro:\*
  - 8: em\_estc:\*
  - 9: gb\_est1:\*
  - 10: gb\_est2:\*
  - 11: gb\_hrc:\*
  - 12: gb\_est3:\*
  - 13: gb\_est4:\*
  - 14: gb\_est5:\*
  - 15: em\_estfun:\*
  - 16: em\_estom:\*
  - 17: em\_gss\_hum:\*
  - 18: em\_gss\_inv:\*
  - 19: em\_gss\_pln:\*
  - 20: em\_gss\_vrt:\*
  - 21: em\_gss\_fun:\*
  - 22: em\_gss\_mam:\*
  - 23: em\_gss\_mus:\*
  - 24: em\_gss\_pro:\*
  - 25: em\_gss\_rod:\*
  - 26: em\_gss\_prg:\*
  - 27: em\_gss\_vrl:\*
  - 28: gb\_gss1:\*
  - 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	19	100.0	721	12	BI099738
2	18	94.7	476	12	BG153958
3	17.4	91.6	182	29	AI753282
4	17.4	91.6	200	9	AUI81058

5	17.4	91.6	269	28	CC069007
6	17.4	91.6	717	14	CA266981
7	17.4	91.6	874	10	BF253672
8	17.4	91.6	881	29	CG172122
9	17.4	91.6	891	28	CC007495
10	17.4	91.6	473	14	CB003069
11	17.4	91.6	571	12	B279225
12	17.4	91.6	551	14	CB003713
13	17.4	91.6	551	14	CB003776
14	17.4	91.6	551	14	CB003990
15	17.4	91.6	551	14	CB003990
16	16.4	86.3	325	10	BE604097
17	16.4	86.3	351	28	BH409827
18	16.4	86.3	402	11	CNS08W6E
19	16.4	86.3	409	13	B0293385
20	16.4	86.3	414	13	B0470363
21	16.4	86.3	417	28	CG179114
22	16.4	86.3	419	20	BF481686
23	16.4	86.3	425	29	CG859216
24	16.4	86.3	430	13	C72864
25	16.4	86.3	440	12	BG465207
26	16.4	86.3	441	13	CA098694
27	16.4	86.3	458	13	C27235
28	16.4	86.3	468	29	CG343674
29	16.4	86.3	476	14	CF019681
30	16.4	86.3	477	13	BU098652
31	16.4	86.3	478	12	BI674934
32	16.4	86.3	500	14	CD231214
33	16.4	86.3	507	13	BU036763
34	16.4	86.3	514	12	BI643448
35	16.4	86.3	521	14	CA233035
36	16.4	86.3	526	14	CD33352
37	16.4	86.3	527	12	BM344683
38	16.4	86.3	531	10	BF481522
39	16.4	86.3	537	13	CA143994
40	16.4	86.3	538	10	BF292414
41	16.4	86.3	544	14	CF771576
42	16.4	86.3	551	14	CA201665
43	16.4	86.3	555	10	AW331233
44	16.4	86.3	558	14	CF266626
45	16.4	86.3	560	12	BG518383
46	16.4	86.3	562	10	BE039189
47	16.4	86.3	562	14	CD425654
48	16.4	86.3	563	10	BE125934
49	16.4	86.3	569	14	CF646593
50	16.4	86.3	570	29	CG106901
51	16.4	86.3	573	10	BF444337
52	16.4	86.3	574	10	AW746004
53	16.4	86.3	579	12	BU465047
54	16.4	86.3	587	12	BI674935
55	16.4	86.3	590	13	CA075992
56	16.4	86.3	592	13	CA076078
57	16.4	86.3	597	13	BO621091
58	16.4	86.3	600	13	CA153745
59	16.4	86.3	601	14	CA755053
60	16.4	86.3	602	28	BH832864
61	16.4	86.3	604	28	BZ775705
62	16.4	86.3	605	12	BU277138
63	16.4	86.3	605	13	CA087427
64	16.4	86.3	605	14	CB639918
65	16.4	86.3	607	10	AW745927
66	16.4	86.3	608	29	CC731048
67	16.4	86.3	615	28	BH832846
68	16.4	86.3	617	9	AI673867
69	16.4	86.3	621	14	CA270256
70	16.4	86.3	622	14	CA188711
71	16.4	86.3	623	14	CF625610
72	16.4	86.3	627	14	CF972352
73	16.4	86.3	633	14	CA252506
74	16.4	86.3	638	14	CA215942
75	16.4	86.3	639	13	BU036764
76	16.4	86.3	643	14	CA220947
77	16.4	86.3	644	28	CC374520

78	16.4	86.3	652	10	BE357618	DGL 21 D0	151	16.4	86.3	959	14	CK254221	CK254221	CK254221	EST737658
79	16.4	86.3	654	28	BH928584	BACPP27-N	152	16.4	86.3	963	14	CK244036	CK244036	CK244036	EST727673
80	16.4	86.3	656	14	CF627049	zmrws05 0	153	16.4	86.3	967	14	CK256446	CK256446	CK256446	EST740083
81	16.4	86.3	663	13	BQ462905	HI02G03F	154	16.4	86.3	975	14	CK272783	CK272783	CK272783	EST718861
82	16.4	86.3	664	29	CG354182	OGXC139TH	155	16.4	86.3	979	14	CK249986	CK249986	CK249986	EST733623
83	16.4	86.3	665	29	CG258824	CG258824	156	16.4	86.3	979	14	CK251027	CK251027	CK251027	EST734664
84	16.4	86.3	669	14	CD883845	CD883845	157	16.4	86.3	982	14	CK254225	CK254225	CK254225	EST737862
85	16.4	86.3	670	9	AU070263	AU070263	158	16.4	86.3	982	14	CK254225	CK254225	CK254225	EST737862
86	16.4	86.3	671	14	CF631174	zmrws48 0	159	16.4	86.3	1413	11	AY103665	AY103665	AY103665	zmrws48 0
87	16.4	86.3	677	14	CA290105	SCAGFL801	160	16.4	86.3	207	12	BM373596	BM373596	BM373596	zmrws48 0
88	16.4	86.3	677	14	CF628006	zmrws05 0	161	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
89	16.4	86.3	683	14	CK253866	CK253866	162	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
90	16.4	86.3	684	29	CG617972	CG617972	163	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
91	16.4	86.3	685	14	CA171503	SCSBSB105	164	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
92	16.4	86.3	685	14	CF627858	zmrws05 0	165	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
93	16.4	86.3	686	13	CA102475	SCRGHR106	166	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
94	16.4	86.3	688	14	CA484842	WH4311 C	167	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
95	16.4	86.3	688	14	CA756082	BR0300340	168	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
96	16.4	86.3	693	10	AW565610	LGL 346 E	169	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
97	16.4	86.3	694	14	CA201745	CA201745	170	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
98	16.4	86.3	694	28	B2962628	PURFIE86TB	171	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
99	16.4	86.3	699	12	BI406624	BI406624	172	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
100	16.4	86.3	699	14	CD218960	CD218960	173	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
101	16.4	86.3	702	9	AU055864	AU055864	174	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
102	16.4	86.3	704	14	CA176030	SCULST102	175	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
103	16.4	86.3	706	14	CD863279	CD863279	176	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
104	16.4	86.3	706	14	CK259264	CK259264	177	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
105	16.4	86.3	712	28	A2130497	OSJNB010	178	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
106	16.4	86.3	714	10	BE357648	DGL 21 D0	179	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
107	16.4	86.3	714	13	CA072793	SCPEAM102	180	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
108	16.4	86.3	717	14	CK243563	CK243563	181	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
109	16.4	86.3	719	14	CB678634	CB678634	182	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
110	16.4	86.3	724	14	CD425969	CD425969	183	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
111	16.4	86.3	729	14	CD881998	CD881998	184	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
112	16.4	86.3	730	13	CA152520	CA152520	185	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
113	16.4	86.3	737	14	CD909820	CD909820	186	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
114	16.4	86.3	743	14	CD430205	ETH1 113 K	187	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
115	16.4	86.3	747	13	BQ801315	BQ801315	188	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
116	16.4	86.3	755	29	CA187912	CA187912	189	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
117	16.4	86.3	778	14	CA215957	CA215957	190	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
118	16.4	86.3	782	14	CA221025	CA221025	191	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
119	16.4	86.3	783	14	CB677195	CB677195	192	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
120	16.4	86.3	795	14	CK252664	CK252664	193	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
121	16.4	86.3	799	28	AQ576888	AQ576888	194	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
122	16.4	86.3	805	13	BQ838243	BQ838243	195	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
123	16.4	86.3	807	14	CA215956	CA215956	196	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
124	16.4	86.3	810	14	CK192897	CK192897	197	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
125	16.4	86.3	813	29	CG731056	CG731056	198	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
126	16.4	86.3	819	29	CG731056	CG731056	199	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
127	16.4	86.3	820	13	CA090915	CA090915	200	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
128	16.4	86.3	828	14	CK254960	CK254960	201	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
129	16.4	86.3	828	28	BZ565381	BZ565381	202	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
130	16.4	86.3	829	29	CG668217	CG668217	203	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
131	16.4	86.3	831	29	CG738045	CG738045	204	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
132	16.4	86.3	831	14	CK251853	CK251853	205	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
133	16.4	86.3	836	14	CB643663	CB643663	206	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
134	16.4	86.3	867	29	CG203456	CG203456	207	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
135	16.4	86.3	870	14	CK233865	CK233865	208	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
136	16.4	86.3	872	29	CG437693	CG437693	209	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
137	16.4	86.3	880	28	BZ636180	BZ636180	210	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
138	16.4	86.3	881	28	CC193260	CC193260	211	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
139	16.4	86.3	889	10	BF127234	BF127234	212	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
140	16.4	86.3	891	14	CK247284	CK247284	213	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
141	16.4	86.3	903	14	CK244035	CK244035	214	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
142	16.4	86.3	910	14	CK259263	CK259263	215	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
143	16.4	86.3	913	14	CK249734	CK249734	216	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
144	16.4	86.3	913	14	CK249734	CK249734	217	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
145	16.4	86.3	917	28	CC375337	CC375337	218	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
146	16.4	86.3	922	13	EQ674430	EQ674430	219	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
147	16.4	86.3	929	13	CC731493	CC731493	220	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
148	16.4	86.3	939	11	AY104963	AY104963	221	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
149	16.4	86.3	950	28	BZ791435	BZ791435	222	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0
150	16.4	86.3	957	14	CK250012	CK250012	223	16.4	86.3	312	29	FR073596	FR073596	FR073596	zmrws48 0

C 224 15.8 83.2 280 14 CK145179 3530.1.16  
C 225 15.8 83.2 283 11 CNS09611  
C 226 15.8 83.2 287 12 BM375241  
C 227 15.8 83.2 290 12 BM642572  
C 228 15.8 83.2 290 13 BQ760616  
C 229 15.8 83.2 291 28 AZ578266  
C 230 15.8 83.2 298 13 BX767139  
C 231 15.8 83.2 299 12 BML138699  
C 232 15.8 83.2 300 9 AJ436424  
C 233 15.8 83.2 301 11 CNS09KLO  
C 234 15.8 83.2 304 11 CNS09F9S  
C 235 15.8 83.2 314 11 CNS09KY3  
C 236 15.8 83.2 318 14 CA670677  
C 237 15.8 83.2 319 13 CA019691  
C 238 15.8 83.2 325 14 D24505  
C 239 15.8 83.2 328 13 CA026320  
C 240 15.8 83.2 329 11 CNS0907Q  
C 241 15.8 83.2 331 9 AJ280053  
C 242 15.8 83.2 332 11 CNS08GAP  
C 243 15.8 83.2 334 12 BM259076  
C 244 15.8 83.2 337 9 AV629675  
C 245 15.8 83.2 337 12 BG607614  
C 246 15.8 83.2 340 11 CNS08CVC  
C 247 15.8 83.2 341 11 CNS08ZJX  
C 248 15.8 83.2 343 13 BX767371  
C 249 15.8 83.2 345 10 BE637925  
C 250 15.8 83.2 347 28 BH583552  
C 251 15.8 83.2 351 11 CNS098F5  
C 252 15.8 83.2 353 28 BH639855  
C 253 15.8 83.2 354 9 AV630085  
C 254 15.8 83.2 354 14 CA643930  
C 255 15.8 83.2 357 11 CNS08LGT  
C 256 15.8 83.2 357 28 BH678798  
C 257 15.8 83.2 358 9 AV630718  
C 258 15.8 83.2 358 11 CNS08G9K  
C 259 15.8 83.2 359 11 CNS09I21  
C 260 15.8 83.2 360 9 AJ474213  
C 261 15.8 83.2 360 9 AV629992  
C 262 15.8 83.2 361 9 AJ281144  
C 263 15.8 83.2 362 29 CNS010F6  
C 264 15.8 83.2 364 14 CD163241  
C 265 15.8 83.2 366 11 CNS090QB  
C 266 15.8 83.2 371 11 CNS08091  
C 267 15.8 83.2 375 13 BX766047  
C 268 15.8 83.2 378 28 AQ902476  
C 269 15.8 83.2 383 9 AU165848  
C 270 15.8 83.2 384 13 BX767488  
C 271 15.8 83.2 384 28 AQ843713  
C 272 15.8 83.2 386 9 AU162306  
C 273 15.8 83.2 387 12 BM368351  
C 274 15.8 83.2 390 12 BM644751  
C 275 15.8 83.2 390 14 CB816855  
C 276 15.8 83.2 390 29 CG765065  
C 277 15.8 83.2 392 12 BJ201699  
C 278 15.8 83.2 393 13 BQ767691  
C 279 15.8 83.2 398 12 BG544751  
C 280 15.8 83.2 400 12 BM598528  
C 281 15.8 83.2 400 13 BX767307  
C 282 15.8 83.2 400 28 AZ919944  
C 283 15.8 83.2 402 9 AJ284757  
C 284 15.8 83.2 403 13 BQ755135  
C 285 15.8 83.2 406 13 BX601762  
C 286 15.8 83.2 406 13 BX767448  
C 287 15.8 83.2 406 13 BX769027  
C 288 15.8 83.2 407 13 BX766546  
C 289 15.8 83.2 410 13 BX768792  
C 290 15.8 83.2 411 11 CNS096RR  
C 291 15.8 83.2 411 14 CB212396  
C 292 15.8 83.2 412 12 BG159499  
C 293 15.8 83.2 412 13 CA002701  
C 294 15.8 83.2 412 13 CA002765  
C 295 15.8 83.2 416 11 CNS09474  
C 296 15.8 83.2 417 11 CNS09GGQ

C 297 15.8 83.2 417 12 BM584771  
C 298 15.8 83.2 419 28 AQ847623  
C 299 15.8 83.2 422 12 BM617375  
C 300 15.8 83.2 422 29 AG211681

ALIGNMENTS

RESULT 1  
BI099738 721 bp mRNA linear EST 26-JUN-2001  
LOCUS 602884916F1 NCI\_COAP\_Kid14 Mus musculus cDNA clone IMAGE:5040602  
DEFINITION 5', mRNA sequence.  
ACCESSION BI099738  
VERSION BI099738.1 GI:14550631  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 721)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Jeffrey E. Green, M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM1111 row: 0 column: 03  
High quality sequence stop: 589.

FEATURES  
source  
1..721  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5040602"  
/lab\_host="PH10B (T1 phage-resistant)"  
/clone\_lib="NCI\_COAP\_Kid14"  
/note="Organ: Kidney; Vector: pCMV-SPORT6; Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Constructed by Life Technologies. Note: this is a NCI\_COAP Library. |"

ORIGIN  
Query Match 100.0%; Score 19; DB 12; Length 721;  
Best Local Similarity 100.0%; Pred. No. 8e+03;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCCGATGACC 19  
|||||  
DB 549 GTCGACGCCGCCGATGACC 567  
|||||

RESULT 2  
BI153958 476 bp mRNA linear EST 05-FEB-2001  
LOCUS 188 LLN01 Lupinus luteus cDNA, mRNA sequence.  
DEFINITION BI153958  
ACCESSION BI153958.1 GI:12665988  
VERSION BI153958.1  
KEYWORDS EST.  
SOURCE Lupinus luteus (yellow lupine)  
ORGANISM Lupinus luteus  
REFERENCE 1 (bases 1 to 476)  
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.

REFERENCE 1 (bases 1 to 476)  
 AUTHORS Podkowinski, J., Kistiel, A., Grabowska, B. and Legocki, A. B.  
 TITLE ESTs from the early nodules library of *Lupinus luteus*  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Podkowinski J  
 Laboratory of Plant Molecular Biology  
 Institute of Bioorganic Chemistry of Polish Academy of Sciences  
 Nostkowskiego 12/14, 61-704 Poznan, Poland  
 Tel: 0048 61 8528919  
 Fax: 0048 61 8520532  
 Email: jantosp@ibch.poznan.pl  
 POLYA=Yes.

## FEATURES

source

1. 476  
 /organism="Lupinus luteus"  
 /mol\_type="mRNA"  
 /cultivar="ventus"  
 /db\_xref="taxon:3873"  
 /tissue\_type="roots with young nodules"  
 /dev\_stage="roots with immature nodules harvested seven  
 days post inoculation with *Bradyrhizobium* sp. WM9  
 (lupinus)"  
 /lab\_host="E. coli strain SOLR"  
 /clone\_lib="L1N01"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from  
 roots with developing nodules harvested seven days post  
 inoculation with *Bradyrhizobium* sp. WM9 (lupinus). The  
 cDNA was directionally ligated into the Uni-ZAP XR vector  
 from Stratagene and packaged using Gigapack III Gold  
 packaging extracts. Plasmids containing cDNA inserts were  
 excised from the recombinant lambda-ZAP phage using  
 Ex-Assist helper phage and propagated in SOLR cells."

## ORIGIN

Query Match 94.7%; Score 18; DB 12; Length 476;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+04; Indels 0; Gaps 0;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19

Db 235 TCGACGCCGCCGATGACC 252  
 |||||  
 |||||

## RESULT 3

AL759282/c  
 LOCUS AL759282 182 bp DNA linear GSS 18-JUN-2002  
 DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-167C03-013362,  
 genomic survey sequence.  
 ACCESSION AL759282  
 VERSION AL759282.1 GI:21497630  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE

1 Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.  
 and Weisshaar, B.  
 TITLE A pipeline for automated high-throughput generation of FSTs  
 (flanking sequence tags) from Arabidopsis thaliana T-DNA  
 transformed lines  
 JOURNAL Unpublished

## REFERENCE

2 Rosso, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weisshaar, B.  
 TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)  
 for flanking sequence tag based reverse genetics

## JOURNAL

Unpublished

## REFERENCE

3 (bases 1 to 182)  
 Strizhov, N., Rosso, M., Li, Y. and Weisshaar, B.

## TITILE

Direct Submission  
 JOURNAL Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

## COMMENT

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
 This sequence is recovered from the left border of the T-DNA. It  
 indicates an insertion within the locus defined by clone t20g20.  
 The sequences are generated at the MPI for Plant Breeding Research  
 in the context of the GABI-Kat project. GABI-Kat is part of the  
 German Plant Genomics program designated 'GABI'. Information on  
 line availability can be found at:  
 http://www.mpiz-koeln.mpg.de/GABI-Kat/.

## FEATURES

source

Location/Qualifiers  
 1. 182

/organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="GK-167C03-013362"  
 /clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
 /note="PCR was performed on DNA from Arabidopsis thaliana  
 plants (Ti) which were transformed with the T-DNA from  
 vector pAC161. The lines contain one or more T-DNA  
 insertions. The DNA fragment(s) resulting from the PCR  
 were directly sequenced to determine the genomic sequence  
 flanking the insertion. Sequences displaying significant  
 similarity to the A. thaliana nuclear genome sequence were  
 processed for submission. T-DNA derived sequences were  
 removed"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 29; Length 182;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+04;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19

Db 112 GTCGACGCCGCCGATGACC 94  
 |||||  
 |||||

## RESULT 4

AU181058  
 LOCUS AU181058 Rice shoot Oryza sativa (japonica cultivar-group) cDNA  
 DEFINITION clone S6021, mRNA sequence.  
 ACCESSION AU181058  
 VERSION AU181058.1 GI:13447564  
 KEYWORDS EST.

## SOURCE

Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoidae; Oryzaceae; Oryza.

## REFERENCE

1 (bases 1 to 200)  
 Sasaki, T. and Yamamoto, K.  
 TITLE Rice cDNA from etiolated shoot (2001)  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Takuji Sasaki

## NATIONAL INSTITUTE OF AGROBIOLOGICAL RESOURCES

Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki

305-8602, Japan  
 Tel: 81-298-38-7441

Fax: 81-298-38-7468

Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/

PROJECT = 'RGP'

## FEATURES

source

Location/Qualifiers  
 1. 200

/organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="S6021"  
 /dev\_stage="Etiolated shoot (8 days old)"  
 /clone\_lib="Rice shoot"  
 /note="Etiolated shoot (8 days old)"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 9; Length 200;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+04;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGACC 19  
 Db 28 GTCGACGCGCGCGATGACC 46  
 |||||  
 |||||

RESULT 5  
 CC069007  
 LOCUS  
 DEFINITION CSU-K33r.42C23.17 CSU-K33r Aedes aegypti genomic clone  
 CSU-K33r.42C23, genomic survey sequence.  
 ACCESSION CC069007  
 VERSION CC069007.1 GI:29907513  
 KEYWORDS GSS.  
 SOURCE Aedes aegypti (yellow fever mosquito)  
 ORGANISM Aedes aegypti  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;  
 Stegomyia.  
 1 (bases 1 to 369)  
 REFERENCE Lotius,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.  
 AUTHORS End sequencing of Aedes aegypti BACS  
 TITLE Unpublished (2003)  
 JOURNAL Other GSSs: CSU-K33r.42C23.SP6  
 COMMENT Contact: Brendan Loftus  
 Department of Eukaryotic Genomics  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-3543  
 Fax: 301-838-0208  
 Email: enta@tigr.org  
 Library was provided by Susan Brown and Dennis Knudson at Colorado  
 State University.  
 Seq primer: 17  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..369  
 /organism="Aedes aegypti"  
 /mol\_type="genomic DNA"  
 /strain="Rexville"  
 /db\_xref="taxon:7159"  
 /clone="CSU-K33r.42C23"  
 /clone\_lib="CSU-K33r"  
 /note="Vector: pBelobAC11; Site\_1: HindIII"

ORIGIN  
 Query Match 91.6%; Score 17.4; DB 28; Length 369;  
 Best Local Similarity 94.7%; Pred. No. 2.6e+04;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGACC 19  
 Db 220 GTCGTCGCGCGCGATGACC 238  
 |||||  
 |||||

RESULT 6  
 CA266981/c  
 LOCUS  
 DEFINITION SCFLB2060F10.g LB2 Saccharum officinarum cDNA clone SCFLB2060F10  
 5', mRNA sequence.  
 ACCESSION CA266981  
 VERSION CA266981.1 GI:35966988  
 KEYWORDS EST.  
 SOURCE Saccharum officinarum  
 ORGANISM Saccharum officinarum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoideae; Andropogoneae; Saccharum.  
 1 (bases 1 to 717)  
 REFERENCE Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

The libraries that made SUCEST  
 Genet. Mol. Biol. 24 (1-4), 1-7 (2001)  
 Contact: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089  
 Email: parruda@unicamp.br  
 Clone distribution: clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
 http://www.bccccenter.fcav.unesp.br  
 Plate: 060 row: F column: 10  
 Seq primer: T7 Promoter Primer.  
 Location/Qualifiers  
 1..717  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCFLB2060F10"  
 /lab\_host="DH10B"  
 /clone\_lib="LB2"  
 /note="Organ: lateral buds from plants adult plants  
 growing in greenhouse; Vector: pSport1; Site\_1: SalI;  
 Site\_2: NotI; An unidirectional cDNA library generated  
 from [lateral buds from plants adult plants growing in  
 greenhouse]. cDNA was prepared from polyA+ mRNA using  
 SuperScript Plasmid System Kit (Invitrogen). The  
 double-strand cDNAs were fractionated in a sepharose  
 Cx-2B 40cm-columns and fragments sizing between 0.8 and  
 1.5 Kb were directionally cloned into the vector. Details  
 of each source of RNA and library construction can be  
 obtained at http://sucest.fad.ic.unicamp.br/public"

ORIGIN  
 Query Match 91.6%; Score 17.4; DB 14; Length 717;  
 Best Local Similarity 94.7%; Pred. No. 2.7e+04;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCGCGCGATGACC 19  
 Db 250 GTCGACGCGCGCGATGACC 232  
 |||||  
 |||||

RESULT 7  
 BF253672/c  
 LOCUS  
 DEFINITION BF253672 874 bp mRNA linear EST 22-OCT-2001  
 HVSMEF0001M19f Hordeum vulgare subsp. vulgare cDNA  
 (St isolated and unstressed) Hordeum vulgare subsp. vulgare cDNA  
 BF253672 GI:13116653  
 ACCESSION BF253672.2  
 VERSION BF253672.2  
 KEYWORDS EST.  
 SOURCE Hordeum vulgare subsp. vulgare  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Poaceae; Triticeae; Hordeum.  
 1 (bases 1 to 874)  
 REFERENCE Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D.,  
 AUTHORS Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W.,  
 Fenton,R.D., Oates,R. and Main,D.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics: Morex unstressed seedling root cDNA library  
 Unpublished (2001)  
 On Nov 16, 2000 this sequence version replaced gi:11182777.  
 JOURNAL Contact: Wing RA  
 COMMENT Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu

Total hg bases = 321  
Seq primer: ATTACCTCTACTAAGG  
High quality sequence stop: 383.  
Location/Qualifiers  
1. .874

## FEATURES

source  
/organism="Hordeum vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Morex"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="HVGMEF0001M19f"  
/tissue\_type="Seedling root"  
/lab\_host="TJC121"  
/clone\_lib="Hordeum vulgare seedling root EST library"

HVCDA0007 (Etiolated and unstressed)"  
/notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water, nystatin and cefotaxime in covered  
crystallization dishes. Five-day old seedling roots were  
then harvested, total RNA was prepared, poly(A) RNA was  
purified, one primary unamplified cDNA library was made,  
and 1 million pfu were in vivo excised to give pBluescript  
SK(-) cDNA phagemids. These steps were performed in the R3  
Close laboratory at the University of California,  
Riverside (Choi, Close, Fenton). Phagemids were plated and  
picked at the Clemson University Genomics Institute (CUGI)  
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
preparations, DNA sequencing and sequence analysis were  
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,  
Oates, Rambo, Main). The sequence has been trimmed to  
remove vector sequence and contains a minimum of 100 bases  
of phred value 20 or above. For more details on library  
preparation and sequence analysis see  
<http://www.genome.clemson.edu/projects/barley>. To order  
this clone see <http://www.genome.clemson.edu/orders> Also  
see Close RJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley Genetics Newsletter 31:29-30.  
(<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 10; Length 874;  
Best Local Similarity 94.7%; Pred. No. 2.8e+04;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCGATGACC 19

Db 180 GTCGACGCCGCGGTGACC 162

## RESULT 8

CG172122 881 bp DNA linear GSS 21-AUG-2003  
PUJAX68TB ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTa0626L16,  
genomic survey sequence.

CG172122 1 GI:34062920

CG172122

GSS.

Zeas

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 881)

Whiteclaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and  
Bennetzen, J.

Maize Genomics Consortium

Unpublished (2003)

Other GSSs: PUJAX68TD

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: sheared ends.

## FEATURES

source  
Location/Qualifiers  
1. .881  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="E73"  
/db\_xref="taxon:4577"  
/clone="ZMMBTa0626L16"  
/clone\_lib="ZM 0.6 1.0 KB"  
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
Cot selected genomic DNA library"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 29; Length 881;  
Best Local Similarity 94.7%; Pred. No. 2.8e+04;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCGATGACC 19

Db 457 GTCGACGCCGCGATGACC 475

## RESULT 9

CC007495/c 891 bp DNA linear GSS 31-MAR-2003  
PUDH81TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTa188M18,  
genomic survey sequence.

CC007495

CC007495

GSS.

Zeas

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 891)

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and  
Bennetzen, J.

Maize Genomics Consortium

Unpublished (2003)

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: sheared ends.

## FEATURES

source  
Location/Qualifiers  
1. .891  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="E73"  
/db\_xref="taxon:4577"  
/clone="ZMMBTa188M18"  
/clone\_lib="ZM 0.6 1.0 KB"  
/notes="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
Cot selected genomic DNA library"

## ORIGIN

Query Match 91.6%; Score 17.4; DB 28; Length 891;  
Best Local Similarity 94.7%; Pred. No. 2.8e+04;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCGATGACC 19

Db 521 GTCGACGCCGCGATGACC 503



RESULT 10  
CB003069  
LOCUS  
DEFINITION CB003069 473 bp mRNA linear EST 10-JAN-2003  
stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera  
CDNA clone VVB024A10 5, mRNA sequence.  
CB003069  
ACCESSION  
VERSION CB003069.1 GI:27590374  
KEYWORDS  
SOURCE  
ORGANISM Vitis vinifera  
Vitis vinifera  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; Vitaceae; Vitis.  
REFERENCE 1 (bases 1 to 473)  
AUTHORS Cramer,G.R. and Cushman,J.C.  
TITLE An expressed sequence tag database for abiotic stressed leaves of  
Vitis vinifera var. Chardonnay  
JOURNAL  
COMMENT Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu  
PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 0 row: A column: 10  
Seq primer: T3 20mer  
High quality sequence stop: 473.  
FEATURES  
source  
Location/Qualifiers  
1..473  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/db\_xref="taxon:29760"  
/clone="VVB024A10"  
/tissue\_type="leaf"  
/dev\_stage="juvenile and adult"  
/clone\_lib="An expressed sequence tag database for abiotic  
stressed leaves of Vitis vinifera var. Chardonnay"  
/note="vector: Lambda Uni-Zap XR, Bluescript SK-; Site 1:  
EcoRI; Site 2: XhoI; Library construction was performed  
according to Stratagene's recommended protocol for the  
Lambda UniZapXR vector and cDNA synthesis kit."  
ORIGIN  
Query Match 89.5%; Score 17; DB 14; Length 473;  
Best Local Similarity 100.0%; Pred.No. 3.6e+04;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCCGATGA 17  
|||||  
Db 143 GTCGACGCCGCCGATGA 159  
RESULT 11  
BJ279225/c  
LOCUS  
DEFINITION BJ279225 571 bp mRNA linear EST 09-APR-2002  
aestivum CDNA clone whrio09 5', mRNA sequence.  
BJ279225  
ACCESSION  
VERSION BJ279225.1 GI:20101874  
KEYWORDS  
SOURCE  
ORGANISM Triticum aestivum (bread wheat)  
Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooidae; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 571)

AUTHORS Ogihara,Y. and Murai,K.  
TITLE Expressed genes in Triticum aestivum  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.  
FEATURES  
source  
Location/Qualifiers  
1..571  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="whrio09"  
/tissue\_type="root"  
/dev\_stage="Feekes' scale 1"  
/clone\_lib="Y. Ogihara unpublished cDNA library, wh\_r"  
ORIGIN  
Query Match 89.5%; Score 17; DB 12; Length 571;  
Best Local Similarity 100.0%; Pred.No. 3.7e+04;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3 CGACGCCGCCGATGACC 19  
|||||  
Db 375 CGACGCCGCCGATGACC 359  
RESULT 12  
CB003713  
LOCUS  
DEFINITION CB003713 651 bp mRNA linear EST 10-JAN-2003  
VVB032C07 134284 An expressed sequence tag database for abiotic  
stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera  
CDNA clone VVB032C07 5, mRNA sequence.  
CB003713  
ACCESSION  
VERSION CB003713.1 GI:27581018  
KEYWORDS  
SOURCE  
ORGANISM Vitis vinifera  
Vitis vinifera  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; Vitaceae; Vitis.  
REFERENCE 1 (bases 1 to 651)  
AUTHORS Cramer,G.R. and Cushman,J.C.  
TITLE An expressed sequence tag database for abiotic stressed leaves of  
Vitis vinifera var. Chardonnay  
JOURNAL Unpublished (2002)  
COMMENT Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu  
PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 0 row: C column: 07  
Seq primer: T3 20mer  
High quality sequence stop: 651.  
FEATURES  
source  
Location/Qualifiers  
1..651  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/db\_xref="taxon:29760"  
/clone="VVB032C07"  
/tissue\_type="leaf"  
/dev\_stage="juvenile and adult"  
/clone\_lib="An expressed sequence tag database for abiotic  
stressed leaves of Vitis vinifera var. Chardonnay"

/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."

## ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 651;  
Best Local Similarity 100.0%; Pred. No. 3.7e+04;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGA 17  
|||||  
DB 143 GTCGACGCCGCCGATGA 159

## RESULT 13

CB003776  
LOCUS  
DEFINITION VVB033A12\_134410 An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVB033A12 5, mRNA sequence.  
ACCESSION CB003776  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Vitis vinifera

CB003776  
EST.  
Vitis vinifera

## REFERENCE

1 (bases 1 to 651)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.

## AUTHORS

Cramer, G.R. and Cushman, J.C.

## TITLE

An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay

## JOURNAL

## COMMENT

Unpublished (2002)  
Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: jcushman@unr.edu

## PCR Primers

FORWARD: T3 20mer

BACKWARD: T7 21mer

Plate: 0 row: A column: 12

Seq primer: T3 20mer

High quality sequence stop: 651.

## FEATURES

## source

1. .651  
Location/Qualifiers  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/db\_xref="taxon:29760"  
/clone="VVB033A12"  
/tissue\_type="leaf"  
/dev\_stage="juvenile and adult"

/clone\_lib="An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay"  
/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."

## ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 651;  
Best Local Similarity 100.0%; Pred. No. 3.7e+04;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGA 17  
|||||  
DB 143 GTCGACGCCGCCGATGA 159

## RESULT 14

## CB003990

## LOCUS

## DEFINITION

CB003990 653 bp mRNA linear EST 10-JAN-2003  
VVB034E11\_134838 An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay Vitis vinifera cDNA clone VVB034E11 5, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## Vitis vinifera

## Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; Vitaceae; Vitis.

## REFERENCE

## AUTHORS

## TITLE

## COMMENT

## JOURNAL

## Unpublished (2002)

## Contact: Cushman JC

## Department of Biochemistry

## University of Nevada

## MS200, Reno, NV 89557-0014, USA

## Tel: 775-784-1918

## Fax: 775-784-1650

## Email: jcushman@unr.edu

## PCR Primers

## FORWARD: T3 20mer

## BACKWARD: T7 21mer

## Plate: 0 row: E column: 11

## Seq primer: T3 20mer

## High quality sequence stop: 653.

## FEATURES

## source

1. .653  
Location/Qualifiers  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/db\_xref="taxon:29760"  
/clone="VVB034E11"  
/tissue\_type="leaf"  
/dev\_stage="juvenile and adult"

/clone\_lib="An expressed sequence tag database for abiotic stressed leaves of Vitis vinifera var. Chardonnay"  
/note="Vector: Lambda Uni-zap XR, Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Library construction was performed according to Stratagene's recommended protocol for the Lambda UniZapXR vector and cDNA synthesis kit."

## ORIGIN

Query Match 89.5%; Score 17; DB 14; Length 653;

Best Local Similarity 100.0%; Pred. No. 3.7e+04;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGA 17  
|||||

DB 143 GTCGACGCCGCCGATGA 159

RESULT 15

BJ284291

LOCUS

DEFINITION

BU284291 Y. Ogiwara unpublished cDNA library, wheat Triticum aestivum cDNA clone wh1009 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Triticum aestivum (bread wheat)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.

1 (bases 1 to 657)

Ogiwara, Y. and Murai, K.

Expressed genes in Triticum aestivum

Unpublished (2002)

COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@gene.nig.ac.jp.

FEATURES  
source  
1. .657  
Location/Qualifiers  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="wrl109"  
/tissue\_type="root"  
/dev\_stage="Feekes' scale 1"  
/clone\_lib="Y. Ogiwara unpublished cDNA library, Wh\_r"

ORIGIN  
Query Match 89.5%; Score 17; DB 12; Length 657;  
Best Local Similarity 100.0%; Pred. No. 3.7e+04;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CGAGCGCCCGCATGACC 19  
|||||  
Db 283 CGAGCGCCCGCATGACC 299

RESULT 16  
BE604097/c  
LOCUS  
DEFINITION  
Triticum aestivum cDNA clone WHE1401-1404\_E17\_E17, mRNA sequence.  
ACCESSION  
BE604097  
VERSION  
BE604097.1 GI:9861367  
KEYWORDS  
EST.  
SOURCE  
Triticum aestivum (bread wheat)  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.  
REFERENCE  
1 (Bases 1 to 325)  
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Han, P.S., Hsia, C.C.,  
Kang, Y., Lazo, G.R., Miller, R., Nguyen H.T., Rausch, C.J.,  
Seaton, C.L., Tong, J.C. and Zhang, D.  
The structure and function of the expressed portion of the wheat  
genomes - Drought stressed leaf cDNA library  
Unpublished (2000)  
Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@w.usda.gov  
Sequence have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: Stratagene SK primer.  
Location/Qualifiers  
1. 325  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="TAM W101"  
/db\_xref="taxon:4565"  
/clone="WHE1401-1404\_E17\_E17"  
/tissue\_type="Leaf"  
/dev\_stage="Full tillering stage"  
/lab\_host="E. coli SOLR"  
/clone\_lib="Wheat drought stressed leaf cDNA library"  
/note="Vector: Lambda Uni-Zap XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were given a gradual  
stress down to 65% and 78% RWC at Texas Tech University  
(D. Zhang in HT Nguyen lab). Total RNA and poly(A) RNA

were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give pluescript phagemids in the Tj Close lab (Choi, Close), at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

ORIGIN  
Query Match 86.3%; Score 16.4; DB 10; Length 325;  
Best Local Similarity 94.4%; Pred. No. 5.6e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCATGAC 18  
|||||  
Db 207 GTCGACGCCCGCATGAC 190

RESULT 17  
BH409827/c  
LOCUS  
DEFINITION  
1007014F04.2EL\_xl 1007 - RescueMu Grid H Zea mays genomic, genomic survey sequence.  
ACCESSION  
BH409827  
VERSION  
BH409827.1 GI:17577672  
KEYWORDS  
GSS.  
SOURCE  
Zea mays  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoidae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 351)  
Walbot, V.  
Maize genomic sequences found using engineered RescueMu transposon  
Unpublished (2001)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1007014 column: 13  
Class: transposon-tagged.  
Location/Qualifiers  
1. 351  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="mixed background W23/Al88/B73"  
/db\_xref="taxon:4577"  
/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH108"  
/clone\_lib="1007 - RescueMu Grid H"  
/note="Organ: leaf; Vector: RescueMu (engineered from pluescript backbone); Site 1: BamHI; Site 2: BglII;  
RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmldb.iastate.edu' and follow the links for 'RescueMu.' Grid H was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH108 cells were transformed and then screened on LB plates with ampicillin."

ORIGIN  
Query Match 86.3%; Score 16.4; DB 28; Length 351;  
Best Local Similarity 94.4%; Pred. No. 5.6e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



**KEYWORDS**  
**SOURCE**  
**ORGANISM**  
**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

**FEATURES**  
**source**

**FEATURES**  
**source**

**ORIGIN**

**Query Match** 86.3%; Score 16.4; DB 10; Length 419;  
**Best Local Similarity** 94.4%; Pred. NO. 5.7e-04;  
**Matches** 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 2 TCGACGCGCGCGATGACC 19  
 |||||  
**DB** 389 TCGACGCGCGCGATGACC 372  
 |||||

**RESULT 23**  
**CG859216**  
**LOCUS**  
**DEFINITION** linear GSS 19-NOV-2003  
**ACCESSION** ZMWBB0268G15f ZMWBB (ECORI) Zea mays subsp. mays genomic clone  
**VERSION** ZMWBB0268G15 5', genomic survey sequence.  
**KEYWORDS** GSS.  
**SOURCE** CG859216.1 GI:38431929  
**ORGANISM** Zea mays subsp. mays (maize)  
**REFERENCE** Zea mays subsp. mays  
**AUTHORS** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
**TITLE** Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
**JOURNAL** clade; Panicoideae; Andropogoneae; Sorghum.  
**COMMENT** 1 (bases 1 to 425)  
 Bhardt,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,  
 Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.  
 Sequencing of the maize genome at FGI (2003c)  
 Unpublished (2003)  
 Contact: Bhardt.A.K.  
 Dr.Joachim Messing's lab  
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers  
 University

190 Frelinghuysen Road, Piscataway, NJ 08854, USA

Tel: 732 445 3801

Fax: 732 445 5735

Email: bharti@aksman.rutgers.edu

Seq primer: T7

Class: BAC ends

High quality sequence start: 133.

Location/Qualifiers

1..425  
/organism="Zea mays subsp. mays"  
/mol\_type="genomic DNA"  
/cultivar="B73"  
/sub\_species="mays"  
/db\_xref="taxon:4578"  
/clone="ZMWBBC0268G15"  
/lab\_host="Z. coli DH10B"  
/clone\_lib="ZMWBBC (EcoRI)"  
/note="Vector: PTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI"

#### ORIGIN

Query Match 86.3%; Score 16.4; DB 29; Length 425;  
Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCGTGACC 19

|||||  
Db 357 TCGACGCCCGCGTGACC 374

#### RESULT 24

##### LOCUS

C72864 430 bp mRNA linear EST 04-APR-2002  
C72864 Rice panicle at flowering stage Oryza sativa (japonica  
cultivar-group) cDNA clone E2385\_1A, mRNA sequence.

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### COMMENT

Sasaki T. and Yamamoto K.  
Rice cDNA from panicle at flowering stage  
Unpublished (1996)  
Contact: Takuji Sasaki  
National Institute of Agrobiological Resources  
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki  
305-8602, Japan  
Tel: 81-298-38-7441  
Fax: 81-298-38-7468  
Email: tsasaki@ab.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/  
PROJECT = "RGP".

#### FEATURES

##### source

1..430  
Location/Qualifiers  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="E2385\_1A"  
/dev\_stage="flowering stage"  
/clone\_lib="Rice panicle at flowering stage"  
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

#### ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 430;  
Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCGTGACC 19

|||||

Db 218 TCGACGCCCGCGTGACC 235

#### RESULT 25

##### LOCUS

BG465207/c 440 bp mRNA linear EST 20-MAR-2001  
BG465207/c EM1.73.B05.g1.A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA  
sequence.

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

BG465207.1 GI:13394183  
Sorghum bicolor (sorghum)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

##### COMMENT

1 (bases 1 to 440)  
Reid, S.P., Cordonnier-Pratt, M.-M., Gingle, A. and Pratt, L.H.  
An EST database from Sorghum: developing embryos  
Unpublished (2000)  
Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions  
below Phred quality 16. The threshold for highest quality sequence  
is 20.  
Seq primer: PolyTwix  
High quality sequence start: 5  
High quality sequence stop: 437  
POLYA=No.

#### FEATURES

##### source

1..440  
Location/Qualifiers  
/organism="Sorghum bicolor"  
/mol\_type="mRNA"  
/db\_xref="taxon:4558"  
/clone\_lib="Embryo 1 (EM1)"  
/note="Organ: Embryos germinated for 24 hr; Vector:  
pBluescript II from lambda Zap II; Site 1: XhoI; Site 2:  
EcoRI; The library was made from poly-A-RNA in the cloning  
vector lambda Zap II. Clones to be sequenced were  
prepared by mass excision."

#### ORIGIN

Query Match 86.3%; Score 16.4; DB 12; Length 440;  
Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCCGCGTGACC 19

|||||

Db 84 TCGACGCCCGCGTGACC 67

#### RESULT 26

##### LOCUS

CA098694 441 bp mRNA linear EST 23-SEP-2003  
SCMCL6060H11.g Cl6 Saccharum officinarum cDNA clone SCMCL6060H11  
5', mRNA sequence.

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

CA098694.1 GI:34952001  
Saccharum officinarum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Saccharum.

##### REFERENCE

##### AUTHORS

##### TITLE

##### JOURNAL

1 (bases 1 to 441)  
Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.  
The libraries that made SUCEST  
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

## COMMENT

Contact: Arruda P  
 Centro de Biologia Molecular e Engenharia Genetica  
 Universidade Estadual de Campinas  
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089  
 Email: parruda@unicamp.br  
 Clone distribution: clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
<http://www.bccccenter.fcav.unesp.br>  
 Tel: 060 row: H column: 11  
 Plate: 060  
 Seq primer: T7 Promoter Primer.  
 Location/Qualifiers  
 1. .441  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCMCL6060H11"  
 /lab\_host="DH10B"  
 /clone\_lib="CJ6"  
 /note="Organ: Pool of sugarcane calli submitted to low  
 (40C) and high (37 C) temperature stress; Vector: pSport1;  
 Site\_1: SalI; Site\_2: NotI; An unidirectional cDNA library  
 generated from [Pool of sugarcane calli submitted to low  
 (40C) and high (37 C) temperature stress]. cDNA was  
 prepared from polyA+ mRNA using SuperScript Plasmid System  
 Kit (Invitrogen). The double-strand cDNAs were  
 fractionated in a sepharose CL-2B 40cm-columns and  
 fragments sizing cloned into the vector. Details of each  
 directionally cloned into the vector. Details of each  
 source of RNA and library construction can be obtained at  
<http://sucest.lad.ic.unicamp.br/public>"

## FEATURES

source

Query Match 86.3%; Score 16.4; DB 13; Length 441;  
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCGACGCGCGCGATGAC 18  
 |||||  
 Db 184 GTCGACGCGCGCGATGAC 167

## ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 441;  
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCGACGCGCGCGATGAC 18  
 |||||  
 Db 184 GTCGACGCGCGCGATGAC 167

## RESULT 27

C27235/c

LOCUS

DEFINITION C27235 Rice callus cDNA Oryza sativa (japonica cultivar-group) cDNA  
 clone C51380\_2A, mRNA sequence.

ACCESSION

C27235

VERSION

EST.

SOURCE

ORGANISM Oryza sativa (japonica cultivar-group)  
 Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Takuji Sasaki  
 National Institute of Agrobiological Resources  
 Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki  
 305-8602, Japan  
 Tel: 81-298-38-7441  
 Fax: 81-298-38-7468  
 Email: tsasaki@abrr.affrc.go.jp, URL: <http://rgp.dna.affrc.go.jp/>  
 PROJECT = "RGP".  
 Location/Qualifiers  
 1. .458  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"

## FEATURES

source

Query Match 86.3%; Score 16.4; DB 29; Length 468;  
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 TCGACGCGCGCGATGACC 19  
 |||||  
 Db 221 TCGACGCGCGCGATGACC 204

/db\_xref="taxon:39947"  
 /clone="C51380\_2A"  
 /tissue\_type="callus"  
 /dev\_stage="callus"  
 /clone\_lib="Rice callus cDNA"

## ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 458;  
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCGACGCGCGCGATGAC 18  
 |||||  
 Db 332 GTCGACGTCGCGATGAC 315

## RESULT 28

CG342674/c

LOCUS

DEFINITION CG342674 468 bp DNA linear GSS 26-AUG-2003  
 OGWAJ86TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0519004,  
 genomic survey sequence.

ACCESSION

CG342674

VERSION

GSS.

KEYWORDS

SOURCE

ORGANISM

Zea mays

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

TIGR

Other\_GSSs: OGWAJ86TV

Contact: Cathy Whitelaw

9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.

## FEATURES

source

Location/Qualifiers  
 1. .468  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBMA0519004"  
 /clone\_lib="ZM 0.7 1.5 KB"  
 /note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

## ORIGIN

Query Match 86.3%; Score 16.4; DB 29; Length 468;  
 Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 TCGACGCGCGCGATGACC 19  
 |||||  
 Db 221 TCGACGCGCGCGATGACC 204

## RESULT 29

CF019681/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

CF019681 476 bp mRNA linear EST 17-JUL-2003  
 QBN21e01.xg QBN Zea mays cDNA clone QBN21e01, mRNA sequence.  
 CF019681  
 CF019681  
 CF019681.1 GI:32914869  
 EST.  
 Zea mays

ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 476)  
REFERENCE Genoplante.  
AUTHORS Genoplante, a major partnership french program in plant genomics  
TITLE Unpublished (2003)  
JOURNAL Contact: Genoplante  
COMMENT Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'genoplante' (<http://www.genoplante.com>  
and <http://genoplante-info.infobiogen.fr>).  
FEATURES Location/Qualifiers  
source 1..476.  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="P2"  
/db\_xref="taxon:4577"  
/clone="QBN21e01"  
/tissue\_type="pedicel, whole kernel"  
/clone\_lib="QBN"

## ORIGIN

Query Match 86.3%; Score 16.4; DB 14; Length 476;  
Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCGACGCCGCCGATGACC 19  
|||||  
Db 434 TCGACGCCGCCGATGACC 417

## RESULT 30

BU098652/2  
LOCUS 946129803.y3 946 - tassal primordium prepared by schmidt lab Zea  
DEFINITION 477 bp mRNA linear EST 29-AUG-2002  
mays cDNA, mRNA sequence.  
ACCESSION BU098652  
VERSION BU098652.1 GI:22546341  
KEYWORDS EST.  
SOURCE Zea mays

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 477)

## REFERENCE

AUTHORS Walbot V.  
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
JOURNAL University  
COMMENT Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: [walbot@stanford.edu](mailto:walbot@stanford.edu)  
Plate: 946129 row: E column: 03.

## FEATURES

source 1..477  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="OH43"  
/db\_xref="taxon:4577"  
/tissue\_type="tassels"  
/dev\_stage="just after the transition from vegetative to  
inflorescence development"  
/lab\_host="XLOLR"  
/clone\_lib="946 - tassal primordium prepared by Schmidt"

lab"  
/note="Organ: tassels; Vector: HybrizAP; Site 1: EcoRI;  
Site 2: XhoI; George Chuck dissected immature tassels  
between 1mm and 3mm. Sharon Stanfield prepared the cDNA  
library in HybrizAP. Sample insert size range was 350 bp  
to 3 Kb with a 1 Kb average."

## ORIGIN

Query Match 86.3%; Score 16.4; DB 13; Length 477;  
Best Local Similarity 94.4%; Pred. No. 5.7e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCGACGCCGCCGATGACC 19  
|||||  
Db 436 TCGACGCCGCCGATGACC 419

Search completed: June 20, 2004, 14:15:09  
Job time : 911.873 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 21.0106 Seconds  
(without alignments)

501.846 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19

Sequence: 1 gtcagccgcgcgatgacc 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PTCUTS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	401	3	US-08-990-823-75
2	19	100.0	401	4	US-09-477-135A-75
3	19	100.0	4403765	3	US-09-103-840A-2
4	19	100.0	4411529	3	US-09-103-840A-1
5	16.4	86.3	1560	4	US-09-598-747-26
6	16.4	86.3	4403765	3	US-09-103-840A-2
7	16.4	86.3	4411529	3	US-09-103-840A-1
8	15.8	83.2	372	4	US-09-252-991A-14184
9	15.8	83.2	459	4	US-09-252-991A-4438
10	15.8	83.2	486	4	US-09-252-991A-1571
11	15.8	83.2	486	4	US-09-252-991A-10164
12	15.8	83.2	528	4	US-09-252-991A-4948
13	15.8	83.2	582	4	US-09-252-991A-1450
14	15.8	83.2	696	4	US-09-252-991A-4312
15	15.8	83.2	705	4	US-09-252-991A-5097
16	15.8	83.2	723	4	US-09-252-991A-4999
17	15.8	83.2	741	4	US-09-252-991A-11040
18	15.8	83.2	861	4	US-09-252-991A-13381
19	15.8	83.2	906	4	US-09-252-991A-9912
20	15.8	83.2	909	4	US-09-252-991A-1620
21	15.8	83.2	921	4	US-09-489-039A-1939
22	15.8	83.2	936	4	US-09-252-991A-4541
23	15.8	83.2	939	4	US-09-252-991A-13966
24	15.8	83.2	951	4	US-09-252-991A-16343
25	15.8	83.2	969	4	US-09-252-991A-15940
26	15.8	83.2	1002	4	US-09-252-991A-11214
27	15.8	83.2	1047	4	US-09-252-991A-13479

28	15.8	83.2	1068	4	US-09-252-991A-1546	Sequence 1546, Ap
29	15.8	83.2	1119	4	US-09-252-991A-16146	Sequence 16146, A
30	15.8	83.2	1902	4	US-09-252-991A-10005	Sequence 10005, A
31	15.8	83.2	2016	4	US-09-252-991A-1449	Sequence 1449, Ap
32	15.8	83.2	2022	4	US-09-252-991A-1573	Sequence 1573, Ap
33	15.8	83.2	2208	4	US-09-252-991A-13998	Sequence 13998, A
34	15.8	83.2	2439	4	US-09-252-991A-14139	Sequence 14139, A
35	15.8	83.2	4293	4	US-09-252-991A-1619	Sequence 1619, Ap
36	15.8	83.2	4359	4	US-09-252-991A-1547	Sequence 1547, Ap
37	15.8	83.2	6449	6	5183745-1	Patent No. 5183745
38	15.8	83.2	6441	4	US-08-669-785-1	Sequence 1, Appl
39	15.8	83.2	6443	6	5183745-5	Patent No. 5183745
40	15.4	81.1	873	4	US-09-252-991A-1761	Sequence 1761, Ap
41	15.4	81.1	972	4	US-09-489-039A-1291	Sequence 1291, Ap
42	15.4	81.1	1032	4	US-09-489-039A-3159	Sequence 3159, Ap
43	15.4	81.1	1191	4	US-09-252-991A-2048	Sequence 2048, Ap
44	15.4	81.1	1647	4	US-09-252-991A-5058	Sequence 5058, Ap
45	15.4	81.1	1928	4	US-09-186-276B-55	Sequence 55, Appl
46	15.4	81.1	1928	4	US-08-842-445-55	Sequence 55, Appl
47	15.4	81.1	1928	4	US-09-186-188B-55	Sequence 55, Appl
48	15.4	81.1	3306	1	US-08-261-206A-71	Sequence 71, Appl
49	15.4	81.1	3546	4	US-09-252-991A-1909	Sequence 1909, Ap
50	15	78.9	77536	4	US-09-410-551B-1	Sequence 1, Appl
51	14.8	77.9	271	4	US-09-313-294A-2856	Sequence 2856, Ap
52	14.8	77.9	282	4	US-09-313-294A-6425	Sequence 6425, Ap
53	14.8	77.9	297	4	US-09-252-991A-5579	Sequence 5579, Ap
54	14.8	77.9	298	4	US-09-313-294A-6043	Sequence 6043, Ap
55	14.8	77.9	336	4	US-09-252-991A-7007	Sequence 7007, Ap
56	14.8	77.9	375	4	US-09-252-991A-4191	Sequence 4191, Ap
57	14.8	77.9	543	4	US-09-489-039A-5141	Sequence 5141, Ap
58	14.8	77.9	558	4	US-09-252-991A-3318	Sequence 3318, Ap
59	14.8	77.9	558	4	US-09-252-991A-12270	Sequence 12270, A
60	14.8	77.9	576	4	US-09-252-991A-2760	Sequence 2760, Ap
61	14.8	77.9	657	4	US-09-252-991A-4339	Sequence 4339, Ap
62	14.8	77.9	738	4	US-09-252-991A-1171	Sequence 1171, Ap
63	14.8	77.9	753	4	US-09-252-991A-1302	Sequence 1302, Ap
64	14.8	77.9	780	4	US-09-489-039A-2858	Sequence 2858, Ap
65	14.8	77.9	801	3	US-09-010-809-11	Sequence 11, Appl
66	14.8	77.9	871	4	US-09-252-991A-15907	Sequence 15907, A
67	14.8	77.9	822	4	US-09-489-039A-5203	Sequence 5203, Ap
68	14.8	77.9	918	4	US-09-252-991A-8161	Sequence 8161, Ap
69	14.8	77.9	933	3	US-09-347-803-21	Sequence 21, Appl
70	14.8	77.9	939	4	US-09-252-991A-4436	Sequence 4436, Ap
71	14.8	77.9	963	4	US-09-252-991A-6979	Sequence 6979, Ap
72	14.8	77.9	966	4	US-09-540-014-27	Sequence 27, Appl
73	14.8	77.9	966	4	US-09-252-991A-2188	Sequence 2188, Ap
74	14.8	77.9	999	4	US-09-252-991A-4170	Sequence 4170, Ap
75	14.8	77.9	999	4	US-09-252-991A-9602	Sequence 9602, Ap
76	14.8	77.9	1011	4	US-09-252-991A-9655	Sequence 9655, Ap
77	14.8	77.9	1044	4	US-09-252-991A-11274	Sequence 11274, A
78	14.8	77.9	1086	4	US-09-252-991A-10259	Sequence 10259, A
79	14.8	77.9	1101	4	US-09-252-991A-10597	Sequence 10597, A
80	14.8	77.9	1122	4	US-09-252-991A-12132	Sequence 12132, A
81	14.8	77.9	1140	4	US-09-252-991A-1313	Sequence 1313, Appl
82	14.8	77.9	1155	4	US-09-252-991A-4515	Sequence 4515, Ap
83	14.8	77.9	1201	2	US-08-169-948B-13	Sequence 13, Appl
84	14.8	77.9	1201	3	US-08-448-873-13	Sequence 13, Appl
85	14.8	77.9	1201	3	US-08-382-452D-13	Sequence 13, Appl
86	14.8	77.9	1201	4	US-08-507-362A-5	Sequence 5, Appl
87	14.8	77.9	1201	4	US-09-316-494A-13	Sequence 13, Appl
88	14.8	77.9	1245	4	US-09-252-991A-2489	Sequence 2489, Ap
89	14.8	77.9	1309	4	US-09-425-578-3	Sequence 3, Appl
90	14.8	77.9	1333	4	US-09-252-991A-9695	Sequence 9695, Ap
91	14.8	77.9	1395	4	US-09-252-991A-3250	Sequence 3250, Ap
92	14.8	77.9	1431	4	US-09-489-039A-2930	Sequence 2930, Ap
93	14.8	77.9	1470	4	US-09-252-991A-2347	Sequence 2347, Ap
94	14.8	77.9	1488	4	US-09-489-039A-943	Sequence 943, App
95	14.8	77.9	1527	1	US-08-380-438-1	Sequence 1, Appl
96	14.8	77.9	1533	4	US-09-252-991A-3013	Sequence 3013, Ap
97	14.8	77.9	1557	4	US-09-252-991A-7033	Sequence 7033, Ap
98	14.8	77.9	1575	4	US-09-489-039A-2196	Sequence 2196, Ap
99	14.8	77.9	1588	2	US-08-389-564B-16	Sequence 16, Appl
100	14.8	77.9	1588	3	US-08-466-047B-16	Sequence 16, Appl

101	14.8	77.9	1662	4	US-09-252-991A-12389	Sequence 12389, A	C 174	14.4	75.8	1335	4	US-09-252-991A-7379	Sequence 7379, Ap
C 102	14.8	77.9	1782	4	US-09-252-991A-15840	Sequence 15840, A	175	14.4	75.8	1386	4	US-09-252-991A-12052	Sequence 12052, A
C 103	14.8	77.9	1848	4	US-09-205-258-188	Sequence 188, App	176	14.4	75.8	1398	4	US-09-252-991A-5150	Sequence 5150, Ap
C 104	14.8	77.9	1866	4	US-09-252-991A-2660	Sequence 2660, Ap	C 177	14.4	75.8	1407	4	US-09-252-991A-5792	Sequence 5792, Ap
C 105	14.8	77.9	1879	2	US-08-403-852D-5	Sequence 5, Appl	C 178	14.4	75.8	1437	4	US-09-252-991A-11925	Sequence 11925, A
C 106	14.8	77.9	1879	3	US-08-510-646B-5	Sequence 5, Appl	C 179	14.4	75.8	1437	4	US-09-252-991A-14884	Sequence 14884, A
C 107	14.8	77.9	1879	3	US-09-231-818-5	Sequence 5, Appl	C 180	14.4	75.8	1482	4	US-09-252-991A-8172	Sequence 8172, Ap
C 108	14.8	77.9	1879	3	US-09-635-359B-5	Sequence 5, Appl	C 181	14.4	75.8	1590	4	US-09-252-991A-246	Sequence 246, App
C 109	14.8	77.9	1929	4	US-09-489-039A-5263	Sequence 5263, Ap	C 182	14.4	75.8	1824	4	US-09-252-991A-5069	Sequence 5069, Ap
C 110	14.8	77.9	2121	4	US-09-489-039A-15877	Sequence 15877, A	C 183	14.4	75.8	1890	4	US-09-252-991A-11961	Sequence 11961, A
C 111	14.8	77.9	2133	4	US-09-489-039A-5273	Sequence 5273, Ap	C 184	14.4	75.8	2000	1	US-08-622-353-1	Sequence 1, Appl
C 112	14.8	77.9	2136	4	US-09-252-991A-13089	Sequence 13089, A	C 185	14.4	75.8	2000	1	US-08-622-352A-1	Sequence 1, Appl
C 113	14.8	77.9	2226	4	US-09-489-039A-822	Sequence 822, App	C 186	14.4	75.8	2000	3	US-08-826-390-1	Sequence 1, Appl
C 114	14.8	77.9	2406	4	US-09-252-991A-12703	Sequence 12703, A	C 187	14.4	75.8	2064	4	US-09-489-039A-2697	Sequence 2697, Ap
C 115	14.8	77.9	2553	4	US-09-252-991A-113	Sequence 113, App	C 188	14.4	75.8	2064	4	US-09-252-991A-13526	Sequence 13526, A
C 116	14.8	77.9	2616	4	US-09-252-991A-2860	Sequence 2860, Ap	C 189	14.4	75.8	2432	4	US-09-252-991A-13933	Sequence 13933, A
C 117	14.8	77.9	2616	4	US-09-252-991A-12851	Sequence 12851, A	C 190	14.4	75.8	2695	4	US-09-252-991A-13826	Sequence 13826, A
C 118	14.8	77.9	2664	4	US-09-252-991A-108	Sequence 108, App	C 191	14.4	75.8	2752	1	US-08-430-925A-3	Sequence 3, Appl
C 119	14.8	77.9	2712	4	US-09-252-991A-12204	Sequence 12204, A	C 192	14.4	75.8	3105	4	US-09-252-991A-4398	Sequence 4398, Ap
C 120	14.8	77.9	2751	4	US-09-252-991A-7066	Sequence 7066, Ap	C 193	14.4	75.8	6909	4	US-09-199-637A-111	Sequence 111, App
C 121	14.8	77.9	3344	1	US-07-718-575-13	Sequence 13, Appl	C 194	14.4	75.8	18318	1	US-08-414-926A-6	Sequence 6, Appl
C 122	14.8	77.9	3344	1	US-08-481-206-13	Sequence 13, Appl	C 195	14.4	75.8	18318	2	US-08-926-922-6	Sequence 6, Appl
C 123	14.8	77.9	3344	2	US-08-486-269A-13	Sequence 13, Appl	C 196	14.4	75.8	18318	3	US-09-253-683-6	Sequence 6, Appl
C 124	14.8	77.9	3978	4	US-09-266-965-19	Sequence 19, Appl	C 197	14.4	75.8	18318	3	US-09-527-657-6	Sequence 6, Appl
C 125	14.8	77.9	4394	1	US-08-095-734-1	Sequence 1, Appl	C 198	14.4	75.8	18318	4	US-09-892-100-6	Sequence 6, Appl
C 126	14.8	77.9	4394	2	US-08-444-623-1	Sequence 1, Appl	C 199	14.4	75.8	35881	4	US-08-311-731A-127	Sequence 127, App
C 127	14.8	77.9	4394	3	US-08-471-869-1	Sequence 1, Appl	C 200	14.4	75.8	35961	4	US-08-311-731A-138	Sequence 138, App
C 128	14.8	77.9	4394	4	US-09-342-563-1	Sequence 1, Appl	C 201	14.4	75.8	536165	4	US-09-214-808-1	Sequence 1, Appl
C 129	14.8	77.9	4394	5	PCT-US94-08267-1	Sequence 1, Appl	C 202	14.2	74.7	25	5	US-09-583-067-1	Sequence 1, Appl
C 130	14.8	77.9	7600	4	US-09-469-211A-1	Sequence 1, Appl	C 203	14.2	74.7	25	5	PCT-US91-06532-5	Sequence 5, Appl
C 131	14.8	77.9	12249	4	US-09-266-965-75	Sequence 75, Appl	C 204	14.2	74.7	51	5	PCT-US91-06532-6	Sequence 6, Appl
C 132	14.8	77.9	18034	4	US-09-266-965-75	Sequence 75, Appl	C 205	14.2	74.7	51	5	US-08-483-533-23	Sequence 23, Appl
C 133	14.8	77.9	18331	4	US-09-266-965-96	Sequence 96, Appl	C 206	14.2	74.7	58	4	US-09-283-471A-23	Sequence 23, Appl
C 134	14.8	77.9	20235	1	US-07-642-734C-3	Sequence 3, Appl	C 207	14.2	74.7	276	4	US-09-313-294A-1799	Sequence 1799, Ap
C 135	14.8	77.9	20235	3	US-08-439-009A-3	Sequence 3, Appl	C 208	14.2	74.7	277	4	US-09-313-294A-5085	Sequence 5085, Ap
C 136	14.8	77.9	49272	1	US-08-614-770A-1	Sequence 1, Appl	C 209	14.2	74.7	278	4	US-09-313-294A-899	Sequence 899, Ap
C 137	14.8	77.9	49377	1	US-08-614-770A-1	Sequence 1, Appl	C 210	14.2	74.7	285	4	US-09-252-991A-11383	Sequence 11383, A
C 138	14.8	77.9	55216	4	US-09-716-865-23	Sequence 23, Appl	C 211	14.2	74.7	289	4	US-09-313-294A-7223	Sequence 7223, Ap
C 139	14.8	77.9	77536	4	US-09-410-551B-1	Sequence 1, Appl	C 212	14.2	74.7	293	4	US-09-313-294A-605	Sequence 605, App
C 140	14.4	75.8	271	4	US-09-313-294A-5134	Sequence 5134, Ap	C 213	14.2	74.7	313	4	US-09-313-294A-6967	Sequence 6967, Ap
C 141	14.4	75.8	278	4	US-09-313-294A-1814	Sequence 1814, Ap	C 214	14.2	74.7	363	4	US-09-252-991A-10416	Sequence 10416, A
C 142	14.4	75.8	289	4	US-09-313-294A-6800	Sequence 6800, Ap	C 215	14.2	74.7	408	4	US-09-252-991A-1365	Sequence 1365, Ap
C 143	14.4	75.8	291	4	US-09-313-294A-4842	Sequence 4842, Ap	C 216	14.2	74.7	408	4	US-09-252-991A-6151	Sequence 6151, Ap
C 144	14.4	75.8	300	4	US-09-313-294A-7553	Sequence 7553, Ap	C 217	14.2	74.7	417	4	US-09-252-991A-12658	Sequence 12658, A
C 145	14.4	75.8	345	4	US-09-252-991A-14621	Sequence 14621, A	C 218	14.2	74.7	426	4	US-09-252-991A-10745	Sequence 10745, A
C 146	14.4	75.8	384	4	US-09-252-991A-4492	Sequence 4492, Ap	C 219	14.2	74.7	435	4	US-09-252-991A-7905	Sequence 7905, Ap
C 147	14.4	75.8	402	4	US-09-252-991A-9829	Sequence 9829, Ap	C 220	14.2	74.7	435	4	US-09-252-991A-14212	Sequence 14212, A
C 148	14.4	75.8	447	4	US-09-252-991A-292	Sequence 292, App	C 221	14.2	74.7	447	4	US-09-252-991A-2898	Sequence 2898, Ap
C 149	14.4	75.8	513	4	US-09-252-991A-11995	Sequence 11995, A	C 222	14.2	74.7	450	4	US-09-252-991A-2601	Sequence 2601, Ap
C 150	14.4	75.8	564	4	US-09-252-991A-11415	Sequence 11415, A	C 223	14.2	74.7	459	4	US-09-427-700-8	Sequence 8, Appl
C 151	14.4	75.8	669	4	US-09-252-991A-5810	Sequence 5810, Ap	C 224	14.2	74.7	471	3	US-09-257-541-6	Sequence 6, Appl
C 152	14.4	75.8	723	4	US-09-252-991A-11457	Sequence 11457, A	C 225	14.2	74.7	474	4	US-09-252-991A-1535	Sequence 1535, Ap
C 153	14.4	75.8	728	4	US-09-252-991A-16380	Sequence 16380, A	C 226	14.2	74.7	474	4	US-09-252-991A-5666	Sequence 5666, Ap
C 154	14.4	75.8	786	4	US-09-252-991A-16489	Sequence 16489, A	C 227	14.2	74.7	483	4	US-09-252-991A-103	Sequence 103, App
C 155	14.4	75.8	789	4	US-09-489-039A-4812	Sequence 4812, Ap	C 228	14.2	74.7	486	4	US-09-621-978-884	Sequence 884, App
C 156	14.4	75.8	807	4	US-09-252-991A-16220	Sequence 16220, A	C 229	14.2	74.7	486	4	US-09-252-991A-7398	Sequence 7398, Ap
C 157	14.4	75.8	886	4	US-09-221-017B-66	Sequence 66, Appl	C 230	14.2	74.7	494	4	US-09-404-879A-196	Sequence 196, App
C 158	14.4	75.8	888	4	US-09-252-991A-11398	Sequence 11398, A	C 231	14.2	74.7	494	4	US-09-338-933-196	Sequence 196, App
C 159	14.4	75.8	894	4	US-09-252-991A-10016	Sequence 10016, A	C 232	14.2	74.7	494	4	US-09-215-681-196	Sequence 196, App
C 160	14.4	75.8	906	4	US-09-252-991A-11438	Sequence 11438, A	C 233	14.2	74.7	494	4	US-09-216-003A-196	Sequence 196, App
C 161	14.4	75.8	933	4	US-09-252-991A-14167	Sequence 14167, A	C 234	14.2	74.7	501	4	US-09-252-991A-5410	Sequence 5410, Ap
C 162	14.4	75.8	1000	4	US-09-050-739-57	Sequence 57, Appl	C 235	14.2	74.7	511	4	US-09-697-367-5	Sequence 5, Appl
C 163	14.4	75.8	1011	4	US-09-252-991A-16115	Sequence 16115, A	C 236	14.2	74.7	516	4	US-09-252-991A-9073	Sequence 9073, Ap
C 164	14.4	75.8	1023	4	US-09-252-991A-16410	Sequence 16410, A	C 237	14.2	74.7	537	4	US-09-252-991A-10311	Sequence 10311, A
C 165	14.4	75.8	1029	4	US-09-489-039A-1491	Sequence 1491, Ap	C 238	14.2	74.7	546	4	US-09-252-991A-8136	Sequence 8136, Ap
C 166	14.4	75.8	1062	4	US-09-252-991A-253	Sequence 253, App	C 239	14.2	74.7	549	4	US-09-252-991A-13068	Sequence 13068, A
C 167	14.4	75.8	1134	4	US-09-252-991A-8078	Sequence 8078, Ap	C 240	14.2	74.7	570	3	US-09-095-855-202	Sequence 202, App
C 168	14.4	75.8	1158	4	US-09-252-991A-14214	Sequence 14214, A	C 241	14.2	74.7	570	4	US-09-205-426-292	Sequence 202, App
C 169	14.4	75.8	1182	4	US-09-252-991A-15009	Sequence 15009, A	C 242	14.2	74.7	570	4	US-09-252-991A-15805	Sequence 15805, A
C 170	14.4	75.8	1200	4	US-09-252-991A-278	Sequence 278, App	C 243	14.2	74.7	572	1	US-07-989-363-1	Sequence 1, Appl
C 171	14.4	75.8	1227	4	US-09-252-991A-14744	Sequence 14744, A	C 244	14.2	74.7	572	1	US-08-264-526-1	Sequence 1, Appl
C 172	14.4	75.8	1242	4	US-09-252-991A-11433	Sequence 11433, A	C 245	14.2	74.7	588	4	US-09-252-991A-2243	Sequence 2243, Ap
C 173	14.4	75.8	1290	4	US-09-252-991A-16182	Sequence 16182, A	C 246	14.2	74.7	594	4	US-09-252-991A-8392	Sequence 8392, Ap



```
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      100.0%; Score 19; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 4219306 GTCGACGCCGCCGATGACC 4219288

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      100.0%; Score 19; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19
Db 4227060 GTCGACGCCGCCGATGACC 4227042

RESULT 5
US-09-598-747-26/c
; Sequence 26, Application US/09598747
; Patent No. 6531648
; GENERAL INFORMATION:
; APPLICANT: Lanahan, Michael B.
; APPLICANT: Desai, Nalini M.
; APPLICANT: Casdaska, Pamela Y.
; TITLE OF INVENTION: GRAIN PROCESSING METHOD AND TRANSGENIC PLANTS USEFUL
; TITLE OF INVENTION: THEREIN
; FILE REFERENCE: A-31383P1
; CURRENT APPLICATION NUMBER: US/09/598,747
; CURRENT FILING DATE: 2000-06-21
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 26
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Oryza sativa
US-09-598-747-26
```

```
Query Match      86.3%; Score 16.4; DB 4; Length 1560;
Best Local Similarity 94.4%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 906 GTCGACGCCGCCGATGAC 889

RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      86.3%; Score 16.4; DB 3; Length 4403765;
Best Local Similarity 94.4%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 277380 GACGACGCCGCCGATGAC 277397

RESULT 7
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      86.3%; Score 16.4; DB 3; Length 4411529;
Best Local Similarity 94.4%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
Db 906 GTCGACGCCGCCGATGAC 889
```

Db 277268 GACGACGCCCGCATGAC 277285  
| |||||  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1571  
; LENGTH: 486  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1571

Query Match 83.2%; Score 15.8; DB 4; Length 486;  
Best Local Similarity 89.5%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCGACGCCCGCATGACC 19  
|||  
Db 311 GTCGACGCCCGCATGACC 329

RESULT 11  
US-09-252-991A-10164/C  
; Sequence 10164, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10164  
; LENGTH: 486  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10164

Query Match 83.2%; Score 15.8; DB 4; Length 486;  
Best Local Similarity 89.5%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCGACGCCCGCATGACC 19  
|||  
Db 123 GTCGACGCCCGCATGACC 105

RESULT 12  
US-09-252-991A-4948  
; Sequence 4948, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4948  
; LENGTH: 528

Db 277268 GACGACGCCCGCATGAC 277285  
| |||||  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 14184  
; LENGTH: 372  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-14184

Query Match 83.2%; Score 15.8; DB 4; Length 372;  
Best Local Similarity 89.5%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCGACGCCCGCATGACC 19  
|||  
Db 33 GTCGACGCCCGCATGACC 51

RESULT 9  
US-09-252-991A-4438/C  
; Sequence 4438, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4438  
; LENGTH: 459  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4438

Query Match 83.2%; Score 15.8; DB 4; Length 459;  
Best Local Similarity 89.5%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCGACGCCCGCATGACC 19  
|||  
Db 148 GTTCCGCCCGCATGACC 130

RESULT 10  
US-09-252-991A-1571  
; Sequence 1571, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4948

Query Match      83.2%; Score 15.8; DB 4; Length 528;
Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 415 GTCGCGCGCGATGACC 433

RESULT 13
US-09-252-991A-1450/c
; Sequence 1450, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1450
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1450

Query Match      83.2%; Score 15.8; DB 4; Length 582;
Best Local Similarity 89.5%; Pred. No. 2e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 74 GTCGACGCGCGATGACC 56

RESULT 14
US-09-252-991A-4312
; Sequence 4312, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4312
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4312

Query Match      83.2%; Score 15.8; DB 4; Length 696;
Best Local Similarity 89.5%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 559 GTTGGCGCGCGATGACC 577

RESULT 15
US-09-252-991A-5097
; Sequence 5097, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5097
; LENGTH: 705
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5097

Query Match      83.2%; Score 15.8; DB 4; Length 705;
Best Local Similarity 89.5%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 453 GCCGACGCCGCGATGACC 471

RESULT 16
US-09-252-991A-4999/c
; Sequence 4999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4999
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4999

Query Match      83.2%; Score 15.8; DB 4; Length 723;
Best Local Similarity 89.5%; Pred. No. 1.9e+02; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCGATGACC 19
Db 384 GTCGCGCGCGATGACC 366

RESULT 17
US-09-252-991A-11040
; Sequence 11040, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
```

```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11040
; LENGTH: 741
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11040

Query Match      83.2%; Score 15.8; DB 4; Length 741;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  GTCGACGCCGCCGATGACC 19
      |||||
DB      178  GTCGACGCCGCCGATACC 196

RESULT 18
US-09-252-991A-13381/c.
; Sequence 13381, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13381
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13381

Query Match      83.2%; Score 15.8; DB 4; Length 861;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  GTCGACGCCGCCGATGACC 19
      |||||
DB      184  GTCGCGCCGCCGATGATC 166

RESULT 19
US-09-252-991A-9912
; Sequence 9912, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9912
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

```

```
; Sequence 4541, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4541
; LENGTH: 936
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4541

Query Match      83.2%; Score 15.8; DB 4; Length 936;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCGACGCCGCCGATGACC 19
Db      384 GTTGGCGCGCCGATGACC 366

RESULT 23
US-09-252-991A-13966
; Sequence 13966, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13966
; LENGTH: 939
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13966

Query Match      83.2%; Score 15.8; DB 4; Length 939;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCGACGCCGCCGATGACC 19
Db      664 GTCGCGCGCGCGATGATC 682

RESULT 24
US-09-252-991A-16343
; Sequence 16343, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```

```
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16343
; LENGTH: 951
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16343

Query Match      83.2%; Score 15.8; DB 4; Length 951;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCGACGCCGCCGATGACC 19
Db      739 GTCGACGCCGCCGCTGATC 757

RESULT 25
US-09-252-991A-15940/c
; Sequence 15940, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15940
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15940

Query Match      83.2%; Score 15.8; DB 4; Length 969;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCGACGCCGCCGATGACC 19
Db      309 GTCGACGCCGCCGCTGATC 291

RESULT 26
US-09-252-991A-11214/c
; Sequence 11214, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11214
; LENGTH: 1002
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11214

Query Match      83.2%; Score 15.8; DB 4; Length 1002;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
```



Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19  
|||||  
Db 705 GTCGACGCCGCCGATACC 587

## RESULT 27

US-09-252-991A-13479/c

; Sequence 13479, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 13479

; LENGTH: 1047

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13479

Query Match 83.2%; Score 15.8; DB 4; Length 1047;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19  
|||||  
Db 441 GTCGACGCCGCCGATGC 423

## RESULT 28

US-09-252-991A-1546

; Sequence 1546, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 1546

; LENGTH: 1068

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-1546

Query Match 83.2%; Score 15.8; DB 4; Length 1068;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19  
|||||  
Db 851 GTCGACGCCGCCGATGCC 869

## RESULT 29

US-09-252-991A-16146/c

; Sequence 16146, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 16146

; LENGTH: 1119

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16146

Query Match 83.2%; Score 15.8; DB 4; Length 1119;

Best Local Similarity 89.5%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19  
|||||  
Db 181 GTCGACGCCGCCGATGC 163

## RESULT 30

US-09-252-991A-10005

; Sequence 10005, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 10005

; LENGTH: 1902

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-10005

Query Match 83.2%; Score 15.8; DB 4; Length 1902;

Best Local Similarity 89.5%; Pred. No. 1.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGACC 19  
|||||  
Db 1836 GTCGACGCCGCCGATGCC 1854

Search completed: June 20, 2004, 11:46:54

Job time : 35.1356 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 107.466 Seconds  
(without alignments)  
809.940 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19

Sequence: 1 gtcgacgcgcgatgacc 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	19	100.0	401	9	US-09-996-634-75	Sequence 75, Appl
2	19	100.0	401	10	US-09-997-182-75	Sequence 75, Appl
3	19	100.0	401	10	US-09-997-181-75	Sequence 75, Appl
4	17.4	91.6	1062	15	US-10-156-761-626	Sequence 626, Appl
5	17.4	91.6	4149	15	US-10-121-988-84	Sequence 84, Appl
6	17.4	91.6	4149	15	US-10-200-582-84	Sequence 84, Appl
7	17.4	91.6	4149	15	US-10-237-551-84	Sequence 84, Appl
8	17.4	91.6	9025608	15	US-10-156-761-1	Sequence 1, Appl
9	17	89.5	25	15	US-10-098-263B-37960	Sequence 37960, A
10	16.4	86.3	246	13	US-10-425-114-22954	Sequence 22954, A
11	16.4	86.3	420	9	US-09-294-093B-1992	Sequence 1992, Ap
12	16.4	86.3	429	16	US-10-260-238-5854	Sequence 5854, Ap
13	16.4	86.3	453	15	US-10-156-761-1762	Sequence 1762, Ap
14	16.4	86.3	646	13	US-10-425-114-15573	Sequence 15573, A

Sequence 23643, A	702	86.3	13	US-10-424-593-23643	Sequence 23643, A
Sequence 2101, Ap	836	86.3	13	US-10-425-114-2101	Sequence 2101, Ap
Sequence 30330, A	836	86.3	13	US-10-425-114-30330	Sequence 30330, A
Sequence 16256, A	880	86.3	13	US-10-425-114-16256	Sequence 16256, A
Sequence 1435, Ap	882	86.3	13	US-10-425-114-1435	Sequence 1435, Ap
Sequence 12308, A	882	86.3	13	US-10-425-114-12308	Sequence 12308, A
Sequence 17287, A	889	86.3	13	US-10-425-114-17287	Sequence 17287, A
Sequence 7159, Ap	893	86.3	13	US-10-425-114-7159	Sequence 7159, Ap
Sequence 120250, A	910	86.3	13	US-10-424-593-120250	Sequence 120250, A
Sequence 3860, Ap	917	86.3	13	US-10-425-114-3860	Sequence 3860, Ap
Sequence 5238, Ap	958	86.3	13	US-10-425-114-5238	Sequence 5238, Ap
Sequence 313, App	1077	86.3	16	US-10-260-238-313	Sequence 313, App
Sequence 5104, Ap	1232	86.3	13	US-10-425-114-5104	Sequence 5104, Ap
Sequence 31945, A	1263	86.3	16	US-10-369-493-31945	Sequence 31945, A
Sequence 1473, Ap	1309	86.3	13	US-10-425-114-1473	Sequence 1473, Ap
Sequence 2580, Ap	1470	86.3	13	US-10-425-114-2580	Sequence 2580, Ap
Sequence 32297, A	1558	86.3	13	US-10-425-114-32297	Sequence 32297, A
Sequence 26, Appl	1560	86.3	15	US-10-306-292-26	Sequence 26, Appl
Sequence 30895, A	1686	86.3	13	US-10-425-114-30895	Sequence 30895, A
Sequence 32140, A	1763	86.3	13	US-10-425-114-32140	Sequence 32140, A
Sequence 32293, A	1935	86.3	16	US-10-369-493-32293	Sequence 32293, A
Sequence 15101, A	1948	86.3	13	US-10-282-122A-15101	Sequence 15101, A
Sequence 4146, Ap	2000	86.3	9	US-09-938-842A-4146	Sequence 4146, Ap
Sequence 4146, Ap	2000	86.3	11	US-09-938-842A-4146	Sequence 4146, Ap
Sequence 32995, A	2357	86.3	13	US-10-425-114-32995	Sequence 32995, A
Sequence 1729, Ap	234	84.2	16	US-09-923-876-1729	Sequence 1729, Ap
Sequence 812, App	234	84.2	11	US-09-923-876-1729	Sequence 812, App
Sequence 13852, A	270	84.2	15	US-10-156-761-812	Sequence 13852, A
Sequence 35, Appl	821	84.2	13	US-10-425-114-13852	Sequence 35, Appl
Sequence 35, Appl	942	84.2	13	US-10-440-503-35	Sequence 35, Appl
Sequence 35, Appl	942	84.2	15	US-10-461-925-35	Sequence 35, Appl
Sequence 35, Appl	942	84.2	15	US-10-146-772-35	Sequence 35, Appl
Sequence 35, Appl	942	84.2	16	US-10-241-742-35	Sequence 35, Appl
Sequence 1970, Ap	1029	84.2	16	US-10-440-523-35	Sequence 1970, Ap
Sequence 7249, Ap	1041	84.2	15	US-10-156-761-1970	Sequence 7249, Ap
Sequence 27555, A	1053	84.2	15	US-10-156-761-7249	Sequence 27555, A
Sequence 45397, A	1381	84.2	16	US-10-369-493-27555	Sequence 45397, A
Sequence 4, Appl	1821	84.2	10	US-09-272-975-4	Sequence 4, Appl
Sequence 4237, A	3732	84.2	16	US-10-369-493-4237	Sequence 4237, A
Sequence 1, Appl	9025608	84.2	15	US-10-156-761-1	Sequence 1, Appl
Sequence 388, App	345	83.2	15	US-10-156-761-388	Sequence 388, App
Sequence 30409, A	477	83.2	13	US-10-282-122A-30409	Sequence 30409, A
Sequence 6597, Ap	600	83.2	15	US-10-156-761-6597	Sequence 6597, Ap
Sequence 2104, Ap	702	83.2	15	US-10-156-761-2104	Sequence 2104, Ap
Sequence 30336, A	738	83.2	13	US-10-282-122A-30336	Sequence 30336, A
Sequence 4028, Ap	742	83.2	16	US-10-260-238-4028	Sequence 4028, Ap
Sequence 40784, A	744	83.2	13	US-10-424-593-40784	Sequence 40784, A
Sequence 2385, Ap	762	83.2	15	US-10-156-761-2385	Sequence 2385, Ap
Sequence 17971, Ap	828	83.2	9	US-09-815-242-7971	Sequence 17971, Ap
Sequence 30566, A	828	83.2	13	US-10-282-122A-30566	Sequence 30566, A
Sequence 1180, Ap	894	83.2	15	US-10-156-761-1180	Sequence 1180, Ap
Sequence 1018, Ap	954	83.2	15	US-10-156-761-1018	Sequence 1018, Ap
Sequence 2402, A	993	83.2	13	US-10-282-122A-32095	Sequence 2402, A
Sequence 5970, Ap	1005	83.2	16	US-10-369-493-24202	Sequence 5970, Ap
Sequence 17694, A	1017	83.2	15	US-10-156-761-5970	Sequence 17694, A
Sequence 34377, A	1053	83.2	16	US-10-369-493-34377	Sequence 34377, A
Sequence 7, Appl	1064	83.2	13	US-10-250-721-7	Sequence 7, Appl
Sequence 7, Appl	1064	83.2	17	US-10-250-824-7	Sequence 7, Appl
Sequence 134, App	1068	83.2	13	US-10-107-431-134	Sequence 134, App
Sequence 42249, A	1107	83.2	16	US-10-369-493-42249	Sequence 42249, A
Sequence 22383, A	1155	83.2	15	US-10-425-114-22383	Sequence 22383, A
Sequence 7033, Ap	1161	83.2	15	US-10-156-761-7033	Sequence 7033, Ap
Sequence 6347, Ap	1182	83.2	13	US-10-156-761-6347	Sequence 6347, Ap
Sequence 22850, A	1294	83.2	15	US-10-425-114-22850	Sequence 22850, A
Sequence 3627, Ap	1350	83.2	15	US-10-156-761-3627	Sequence 3627, Ap
Sequence 32374, A	1366	83.2	16	US-10-369-493-32374	Sequence 32374, A
Sequence 35452, A	1386	83.2	16	US-10-369-493-35452	Sequence 35452, A
Sequence 45481, A	1416	83.2	16	US-10-369-493-45481	Sequence 45481, A
Sequence 44445, A	1425	83.2	15	US-10-369-493-44445	Sequence 44445, A
Sequence 40480, A	1440	83.2	16	US-10-369-493-40480	Sequence 40480, A
Sequence 305, App	1467	83.2	9	US-09-738-626-305	Sequence 305, App

c 88	15.8	83.2	1537	9	US-09-974-300-2395	Sequence 2235, Ap	c 161	15.4	81.1	1496	13	US-10-425-114-2395	Sequence 2395, Ap
c 89	15.8	83.2	1596	16	US-10-369-493-38760	Sequence 38760, A	c 162	15.4	81.1	1500	13	US-10-425-114-2336	Sequence 2236, Ap
c 90	15.8	83.2	1602	16	US-10-369-493-44749	Sequence 44749, A	c 163	15.4	81.1	1513	13	US-10-282-122A-25918	Sequence 25918, A
c 91	15.8	83.2	1620	15	US-10-156-761-6900	Sequence 6900, Ap	c 164	15.4	81.1	1516	13	US-10-425-114-464	Sequence 464, Ap
c 92	15.8	83.2	1713	9	US-09-815-243-7853	Sequence 7853, Ap	c 165	15.4	81.1	1521	15	US-10-156-761-6438	Sequence 6438, Ap
c 93	15.8	83.2	1713	13	US-10-282-122A-30361	Sequence 30361, A	c 166	15.4	81.1	1538	13	US-10-425-114-3634	Sequence 3634, Ap
c 94	15.8	83.2	1772	13	US-10-425-114-15877	Sequence 15877, A	c 167	15.4	81.1	1541	13	US-10-425-114-2706	Sequence 2706, Ap
c 95	15.8	83.2	1791	16	US-10-369-493-41699	Sequence 41699, A	c 168	15.4	81.1	1542	13	US-10-425-114-23170	Sequence 2170, A
c 96	15.8	83.2	1795	16	US-10-260-238-906	Sequence 906, Ap	c 169	15.4	81.1	1543	13	US-10-425-114-25007	Sequence 25007, A
c 97	15.8	83.2	2280	16	US-10-369-493-42316	Sequence 42316, A	c 170	15.4	81.1	1545	13	US-10-425-114-28405	Sequence 28405, A
c 98	15.8	83.2	2373	13	US-10-389-647-288	Sequence 288, Ap	c 171	15.4	81.1	1548	13	US-10-425-114-13947	Sequence 13947, A
c 99	15.8	83.2	2385	16	US-10-369-493-31913	Sequence 31913, A	c 172	15.4	81.1	1552	13	US-10-425-114-27282	Sequence 27282, A
c 100	15.8	83.2	2295	13	US-10-282-122A-14998	Sequence 14998, A	c 173	15.4	81.1	1553	13	US-10-425-114-23112	Sequence 23112, A
c 101	15.8	83.2	2531	16	US-10-369-493-44258	Sequence 44258, A	c 174	15.4	81.1	1554	13	US-10-425-114-2786	Sequence 26786, A
c 102	15.8	83.2	2712	16	US-10-369-493-40442	Sequence 40442, A	c 175	15.4	81.1	1557	13	US-10-425-114-27441	Sequence 27441, A
c 103	15.8	83.2	2920	13	US-10-282-122A-19428	Sequence 19428, A	c 176	15.4	81.1	1560	13	US-10-425-114-27562	Sequence 27562, A
c 104	15.8	83.2	3297	13	US-10-282-122A-30362	Sequence 30362, A	c 177	15.4	81.1	1563	13	US-10-425-114-27611	Sequence 27611, A
c 105	15.8	83.2	3897	13	US-10-282-122A-31557	Sequence 31557, A	c 178	15.4	81.1	1565	13	US-10-425-114-26626	Sequence 26626, A
c 106	15.8	83.2	4125	15	US-10-121-988-76	Sequence 76, Appl	c 179	15.4	81.1	1569	13	US-10-425-114-26725	Sequence 26725, A
c 107	15.8	83.2	4125	15	US-10-121-988-149	Sequence 149, Ap	c 180	15.4	81.1	1573	13	US-10-425-114-27314	Sequence 27314, A
c 108	15.8	83.2	4125	15	US-10-200-562-76	Sequence 76, Appl	c 181	15.4	81.1	1583	13	US-10-425-114-27861	Sequence 27861, A
c 109	15.8	83.2	4125	15	US-10-200-562-149	Sequence 149, Ap	c 182	15.4	81.1	1586	13	US-10-425-114-28803	Sequence 28803, A
c 110	15.8	83.2	4125	15	US-10-200-562-210	Sequence 210, Ap	c 183	15.4	81.1	1596	13	US-10-425-114-27268	Sequence 27268, A
c 111	15.8	83.2	4125	15	US-10-237-551-76	Sequence 76, Appl	c 184	15.4	81.1	1637	13	US-10-425-114-1064	Sequence 1064, Ap
c 112	15.8	83.2	4125	15	US-10-237-551-149	Sequence 149, Ap	c 185	15.4	81.1	1701	13	US-10-425-114-669	Sequence 669, Ap
c 113	15.8	83.2	4125	15	US-10-237-551-210	Sequence 210, Ap	c 186	15.4	81.1	1713	9	US-09-738-626-1004	Sequence 1004, Ap
c 114	15.8	83.2	4443	13	US-10-156-761-3260	Sequence 3260, Ap	c 187	15.4	81.1	1716	13	US-10-425-114-25307	Sequence 25307, A
c 115	15.8	83.2	5811	13	US-10-152-886-2	Sequence 2, Appl	c 188	15.4	81.1	1773	16	US-10-369-493-24296	Sequence 24296, A
c 116	15.8	83.2	6442	9	US-09-950-335A-11	Sequence 11, Appl	c 189	15.4	81.1	1881	16	US-10-369-493-35626	Sequence 35626, A
c 117	15.8	83.2	8009	11	US-09-983-658-34	Sequence 34, Appl	c 190	15.4	81.1	1923	9	US-09-938-842A-2142	Sequence 2142, Ap
c 118	15.8	83.2	8009	11	US-09-983-687-34	Sequence 34, Appl	c 191	15.4	81.1	1923	11	US-09-938-842A-2142	Sequence 2142, Ap
c 119	15.8	83.2	11238	9	US-10-205-032-15	Sequence 15, Appl	c 192	15.4	81.1	1928	13	US-10-283-007-55	Sequence 55, Appl
c 120	15.8	83.2	30365	9	US-09-825-414-1	Sequence 1, Appl	c 193	15.4	81.1	1942	13	US-10-283-122A-14251	Sequence 14251, A
c 121	15.8	83.2	45055	13	US-10-107-431-377	Sequence 277, Ap	c 194	15.4	81.1	2070	13	US-10-087-192-620	Sequence 620, Ap
c 122	15.8	83.2	45191	15	US-10-080-170-549	Sequence 649, Ap	c 195	15.4	81.1	2408	15	US-10-156-761-3357	Sequence 3357, Ap
c 123	15.8	83.2	60196	15	US-10-205-032-1	Sequence 1, Appl	c 196	15.4	81.1	3201	13	US-10-425-114-31255	Sequence 31255, A
c 124	15.8	83.2	82993	15	US-10-080-170-545	Sequence 645, Ap	c 197	15.4	81.1	8391	9	US-09-870-759-113	Sequence 113, Ap
c 125	15.8	83.2	154746	10	US-09-827-688-8	Sequence 8, Appl	c 198	15.4	81.1	8391	10	US-09-751-708A-113	Sequence 113, Ap
c 126	15.8	83.2	3309400	9	US-09-738-626-1	Sequence 1, Appl	c 199	15.4	81.1	9210	9	US-09-712-363-100	Sequence 100, Ap
c 127	15.4	81.1	25	15	US-10-098-626B-37959	Sequence 37959, A	c 200	15.4	81.1	26650	13	US-10-087-192-619	Sequence 619, Ap
c 128	15.4	81.1	212	9	US-09-923-876-2455	Sequence 2455, Ap	c 201	15.4	81.1	3309400	9	US-09-738-626-1	Sequence 5, Appl
c 129	15.4	81.1	212	11	US-09-923-876-2455	Sequence 2455, Ap	c 202	15	78.9	482	10	US-09-918-995-1434	Sequence 35187, A
c 130	15.4	81.1	258	16	US-10-259-194A-612	Sequence 612, Ap	c 203	15	78.9	695	13	US-10-425-114-12513	Sequence 38162, A
c 131	15.4	81.1	258	16	US-10-260-238-5499	Sequence 5499, Ap	c 204	15	78.9	1017	15	US-10-156-761-4643	Sequence 38793, A
c 132	15.4	81.1	342	13	US-10-282-122A-18017	Sequence 18017, A	c 205	15	78.9	1189	9	US-09-900-237-5	Sequence 33966, A
c 133	15.4	81.1	400	13	US-10-062-727-385	Sequence 385, Ap	c 206	15	78.9	1230	16	US-10-369-493-38793	Sequence 27108, A
c 134	15.4	81.1	491	13	US-10-425-114-6800	Sequence 6800, Ap	c 207	15	78.9	1538	16	US-10-369-493-38793	Sequence 29125, A
c 135	15.4	81.1	552	16	US-10-374-780A-1879	Sequence 1879, Ap	c 208	15	78.9	1533	16	US-10-369-493-38793	Sequence 36845, A
c 136	15.4	81.1	558	15	US-10-156-761-3377	Sequence 3377, Ap	c 209	15	78.9	1561	13	US-10-425-114-33596	Sequence 21556, Ap
c 137	15.4	81.1	640	13	US-10-425-114-17063	Sequence 17063, A	c 210	15	78.9	1570	16	US-10-369-493-27108	Sequence 27108, A
c 138	15.4	81.1	688	13	US-10-425-114-4767	Sequence 4767, Ap	c 211	15	78.9	1623	15	US-10-156-761-1556	Sequence 29125, A
c 139	15.4	81.1	729	16	US-10-369-493-28359	Sequence 28359, A	c 212	15	78.9	1632	13	US-10-282-122A-36845	Sequence 36845, A
c 140	15.4	81.1	732	13	US-10-282-122A-13021	Sequence 13021, A	c 213	15	78.9	1632	13	US-10-282-122A-36845	Sequence 36845, A
c 141	15.4	81.1	735	13	US-10-282-122A-25936	Sequence 25936, A	c 214	15	78.9	1817	13	US-10-282-122A-36845	Sequence 36845, A
c 142	15.4	81.1	743	13	US-10-425-114-18343	Sequence 18343, A	c 215	15	78.9	1993	13	US-10-282-122A-38805	Sequence 38805, A
c 143	15.4	81.1	825	13	US-10-282-122A-33115	Sequence 33115, A	c 216	15	78.9	1993	13	US-10-282-122A-39208	Sequence 39208, A
c 144	15.4	81.1	921	13	US-10-425-114-9938	Sequence 9938, Ap	c 217	15	78.9	2205	15	US-10-156-761-4155	Sequence 4155, Ap
c 145	15.4	81.1	945	13	US-10-282-122A-23723	Sequence 23723, A	c 218	15	78.9	2351	13	US-10-425-114-2510	Sequence 1241, Ap
c 146	15.4	81.1	981	13	US-10-282-122A-37435	Sequence 37435, A	c 219	15	78.9	2664	15	US-10-156-761-1241	Sequence 1241, Ap
c 147	15.4	81.1	999	13	US-10-282-122A-38813	Sequence 38813, A	c 220	15	78.9	2687	13	US-10-425-114-833	Sequence 833, Ap
c 148	15.4	81.1	1035	13	US-10-282-122A-39319	Sequence 39319, A	c 221	15	78.9	3028	15	US-10-209-059-29	Sequence 29, Appl
c 149	15.4	81.1	1038	16	US-10-369-493-27245	Sequence 27245, A	c 222	15	78.9	3470	13	US-10-627-132-25	Sequence 25, Appl
c 150	15.4	81.1	1060	13	US-10-425-114-9027	Sequence 9027, Ap	c 223	15	78.9	3470	13	US-10-627-132-25	Sequence 25, Appl
c 151	15.4	81.1	1088	13	US-10-425-114-7065	Sequence 7065, Ap	c 224	15	78.9	3470	15	US-10-209-059-25	Sequence 25, Appl
c 152	15.4	81.1	1168	13	US-10-424-599-16495	Sequence 16495, A	c 225	15	78.9	3589	13	US-10-425-114-24930	Sequence 24930, A
c 153	15.4	81.1	1176	9	US-09-815-242-4068	Sequence 4068, Ap	c 226	15	78.9	77536	10	US-09-940-0316B-1	Sequence 4795, Ap
c 154	15.4	81.1	1197	13	US-10-282-122A-7259	Sequence 7259, Ap	c 227	14.8	77.9	120	9	US-09-294-093B-4795	Sequence 5003, Ap
c 155	15.4	81.1	1197	9	US-09-713-363-14	Sequence 14, Appl	c 228	14.8	77.9	282	15	US-10-156-761-5003	Sequence 8894, A
c 156	15.4	81.1	1265	9	US-09-974-300-2324	Sequence 2324, Ap	c 229	14.8	77.9	293	13	US-10-424-599-8894	Sequence 8894, A
c 157	15.4	81.1	1281	16	US-10-369-493-40595	Sequence 40595, A	c 230	14.8	77.9	315	13	US-10-062-727-848	Sequence 848, Ap
c 158	15.4	81.1	1419	16	US-10-369-493-34306	Sequence 34306, A	c 231	14.8	77.9	318	11	US-09-864-408A-2373	Sequence 2373, Ap
c 159	15.4	81.1	1431	13	US-10-425-114-26383	Sequence 26383, A	c 232	14.8	77.9	394	13	US-10-425-114-15980	Sequence 15980, A
c 160	15.4	81.1	1466	13	US-10-425-114-3280	Sequence 3280, Ap	c 233	14.8	77.9	404	13	US-10-424-599-107233	Sequence 107233, A

US-09-996-634-75  
; Sequence 75, Application US/09996634  
; Patent No. US20020172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 61260  
; CURRENT APPLICATION NUMBER: US/09/996,634  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/447,135  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08/990,823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 75  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(401)  
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.  
US-09-996-634-75  
Query Match 100.0%; Score 19; DB 9; Length 401;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cyt 1 GTCGACGCCCGCATGACC 19  
Db 12 GTCGACGCCCGCATGACC 30  
RESULT 2  
US-09-997-182-75  
; Sequence 75, Application US/09997182  
; Publication No. US20030049263A1  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 61258  
; CURRENT APPLICATION NUMBER: US/09/997,182  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/447,135  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08/990,823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 75  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(401)  
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.  
US-09-997-182-75  
Query Match 100.0%; Score 19; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 9, Appli  
Sequence 29773, A  
Sequence 14472, A  
Sequence 1401, A  
Sequence 14374, A  
Sequence 28827, A  
Sequence 29519, A  
Sequence 3959, A  
Sequence 139738, A  
Sequence 12112, A  
Sequence 14241, A  
Sequence 34345, A  
Sequence 23407, A  
Sequence 145, App  
Sequence 12522, A  
Sequence 183436, A  
Sequence 28803, A  
Sequence 57549, A  
Sequence 201, App  
Sequence 6930, App  
Sequence 1424, App  
Sequence 1267, App  
Sequence 186, App  
Sequence 1995, App  
Sequence 6076, App  
Sequence 39580, A  
Sequence 39593, A  
Sequence 31843, A  
Sequence 29018, A  
Sequence 7843, App  
Sequence 28785, A  
Sequence 30338, A  
Sequence 39209, A  
Sequence 42344, A  
Sequence 2612, App  
Sequence 5022, App  
Sequence 30532, A  
Sequence 118913, A  
Sequence 57527, A  
Sequence 31124, A  
Sequence 16061, A  
Sequence 47111, A  
Sequence 23640, A  
Sequence 227, App  
Sequence 27685, A  
Sequence 26536, A  
Sequence 1426, App  
Sequence 1269, App  
Sequence 28738, A  
Sequence 23637, A  
Sequence 23841, A  
Sequence 23192, A  
Sequence 5117, App  
Sequence 47104, A  
Sequence 30966, A  
Sequence 27621, A  
Sequence 11189, A  
Sequence 329, App  
Sequence 32531, A  
Sequence 31794, A  
Sequence 28207, A  
Sequence 14278, A  
Sequence 5167, App  
Sequence 1027, App  
Sequence 1609, App  
Sequence 21, Appli  
US-10-125-258-9  
US-10-282-122A-29773  
US-10-425-114-14472  
US-10-282-122A-11401  
US-10-282-122A-14374  
US-10-282-122A-28827  
US-10-282-122A-29519  
US-09-734-569-39  
US-10-424-599-139738  
US-10-425-114-12112  
US-10-425-114-14241  
US-10-424-599-34345  
US-10-425-114-23407  
US-10-425-114-145  
US-10-425-114-12522  
US-10-027-632-183436  
US-10-027-632-28803  
US-10-282-122A-28803  
US-10-424-599-57549  
US-10-389-647-201  
US-10-156-761-6930  
US-10-412-699B-1424  
US-10-374-780A-1267  
US-10-260-238-186  
US-10-156-761-1995  
US-10-156-761-6076  
US-10-369-493-39580  
US-10-369-493-39593  
US-10-282-122A-31843  
US-10-425-114-29018  
US-09-815-242-7843  
US-10-282-122A-28785  
US-10-282-122A-30338  
US-10-369-493-39209  
US-10-369-493-42344  
US-10-156-761-2612  
US-10-425-114-5022  
US-10-424-599-30532  
US-10-424-599-118913  
US-10-424-599-57527  
US-10-425-114-31124  
US-10-425-114-16061  
US-10-369-493-47111  
US-10-425-114-23640  
US-09-738-626-227  
US-10-425-114-27685  
US-10-282-122A-26536  
US-10-412-699B-1426  
US-10-374-780A-1269  
US-10-282-122A-28738  
US-10-425-114-23637  
US-10-425-114-23841  
US-10-425-114-23192  
US-10-156-761-5117  
US-10-369-493-47104  
US-10-369-493-30966  
US-10-369-493-27621  
US-10-282-122A-11189  
US-09-974-300-329  
US-10-424-599-32531  
US-10-425-114-31794  
US-10-369-493-28207  
US-10-425-114-14278  
US-10-156-761-5167  
US-10-425-114-1027  
US-10-425-114-1609  
US-09-882-691-21  
ALIGNMENTS  
RESULT 1

QY 1 GTCGACGCCCGCGATGACC 19  
Db 12 GTCGACGCCCGCGATGACC 30

RESULT 3  
US-09-997-181-75  
; Sequence 75, Application US/09997181  
; Publication No. US20030049269A1  
; GENERAL INFORMATION:  
; APPLICANT: Naro, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; FILE REFERENCE: 61257  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 75  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; NAME/KEY: misc feature  
; LOCATION: (1)..(401)  
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.

US-09-997-181-75

Query Match 100.0%; Score 19; DB 10; Length 401;  
Best Local Similarity 100.0%; Pred. No. 8.5;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCGATGACC 19  
Db 12 GTCGACGCCCGCGATGACC 30

RESULT 4  
US-10-156-761-626/c  
; Sequence 626, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 626  
; LENGTH: 1062  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; NAME/KEY: CDS

; LOCATION: (1)..(1062)  
US-10-156-761-626

Query Match 91.6%; Score 17.4; DB 15; Length 1062;  
Best Local Similarity 94.7%; Pred. No. 43;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCGATGACC 19  
Db 36 GTCGACGCCCGCGATGACC 18

RESULT 5  
US-10-121-988-84  
; Sequence 84, Application US/10121988  
; Publication No. US20030068327A1  
; GENERAL INFORMATION:  
; APPLICANT: Hosken, Nancy Ann  
; APPLICANT: McGowan, Patrick  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: Mossman, Sally P.  
; APPLICANT: Evans, Lawrence S.  
; APPLICANT: Swanson, Ryan M.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; FILE REFERENCE: 210121.538C1  
; CURRENT FILING DATE: 2002-04-11  
; CURRENT FILING DATE: 2002-04-11  
; NUMBER OF SEQ ID NOS: 183  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 84  
; LENGTH: 4149  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.

US-10-121-988-84

Query Match 91.6%; Score 17.4; DB 15; Length 4149;  
Best Local Similarity 94.7%; Pred. No. 36;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCGATGACC 19  
Db 2395 GTCGACGCCCGCGATGACC 2413

RESULT 6  
US-10-200-562-84  
; Sequence 84, Application US/10200562  
; Publication No. US20030165819A1  
; GENERAL INFORMATION:  
; APPLICANT: McGowan, Patrick  
; APPLICANT: Hosken, Nancy A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; FILE REFERENCE: 210121.538C2  
; CURRENT FILING DATE: 2002-07-19  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 84  
; LENGTH: 4149  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; OTHER INFORMATION: n is a, c, g, or t/u; w is t/u or a.

US-10-200-562-84

Query Match 91.6%; Score 17.4; DB 15; Length 4149;  
Best Local Similarity 94.7%; Pred. No. 36;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCCGCGATGACC 19  
Db 2395 GTCGACGCCCGCGATGACC 2413

## RESULT 7

US-10-237-551-84  
; Sequence 84, Application US/10237551  
; Publication No. US20030165820A1  
; GENERAL INFORMATION:  
; APPLICANT: Day, Craig H.  
; APPLICANT: Hosken, Nancy A.  
; APPLICANT: Parsons, Joseph M.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND  
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION  
; FILE REFERENCE: 210121.538C3  
; CURRENT APPLICATION NUMBER: US/10/237,551  
; CURRENT FILING DATE: 2002-09-06  
; NUMBER OF SEQ ID NOS: 254  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 84  
; LENGTH: 4149  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-237-551-84

Query Match 91.6%; Score 17.4; DB 15; Length 4149;  
Best Local Similarity 94.7%; Pred. No. 36;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19

DB 2395 GTCGACGCCGCCGATGACC 2413

## RESULT 8

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (418715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 91.6%; Score 17.4; DB 15; Length 9025608;  
Best Local Similarity 94.7%; Pred. No. 13;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19

DB 788283 GTCGACGCCGCCGATGACC 788265

## RESULT 9

US-10-098-263B-37960  
; Sequence 37960, Application US/10098263B  
; Publication No. US20030104410A1  
; GENERAL INFORMATION:  
; APPLICANT: Mittman, Michael  
; TITLE OF INVENTION: Human Microarray  
; FILE REFERENCE: 3118.1  
; CURRENT APPLICATION NUMBER: US/10/098,263B  
; CURRENT FILING DATE: 2003-01-08  
; PRIOR APPLICATION NUMBER: 60/276,759  
; PRIOR FILING DATE: 2001-03-16  
; NUMBER OF SEQ ID NOS: 131066  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 37960  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-10-098-263B-37960

Query Match 89.5%; Score 17; DB 15; Length 25;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGA 17

DB 9 GTCGACGCCGCCGATGA 25

## RESULT 10

US-10-425-114-22954/c  
; Sequence 22954, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 22954  
; LENGTH: 246  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3592-081-C10\_FLI  
US-10-425-114-22954

Query Match 86.3%; Score 16.4; DB 13; Length 246;  
Best Local Similarity 94.4%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19

DB 105 TCGACGCCGCCGATGACC 88

## RESULT 11

US-09-294-093B-1992/c  
; Sequence 1992, Application US/09294093B  
; Patent No. US20010051335A1  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath, V.  
; APPLICANT: Ito, Laura, Y.  
; APPLICANT: Sherman, Bradley, K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL  
; FILE REFERENCE: PL-0009 US  
; CURRENT APPLICATION NUMBER: US/09/294,093B

; CURRENT FILING DATE: 1999-04-16  
; PRIOR APPLICATION NUMBER: 60/082,567  
; PRIOR FILING DATE: April 21, 1998  
; NUMBER OF SEQ ID NOS: 6207  
; SOFTWARE: PERL Program  
; SEQ ID NO 1992  
; LENGTH: 270  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700345337H1  
; NAME/KEY: unsure  
; LOCATION: 68  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-294-093B-1992

Query Match 86.3%; Score 16.4; DB 9; Length 270;  
Best Local Similarity 94.4%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18  
|||||  
Db 87 GTCGACGCCGCCGATGAC 70

RESULT 12  
US-10-260-238-5854/c  
; Sequence 5854, Application US/10260238  
; Publication No. US20040016025A1  
; GENERAL INFORMATION:  
; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Rumiya  
; APPLICANT: Kreps, Joel  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
; FILE REFERENCE: 60111-NP  
; CURRENT APPLICATION NUMBER: US/10/260,238  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,448  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 6077  
; SEQ ID NO 5854  
; LENGTH: 429  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-10-260-238-5854

Query Match 86.3%; Score 16.4; DB 16; Length 429;  
Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19  
|||||  
Db 298 TCGACGCCGCCGATGACC 281

RESULT 13  
US-10-156-761-1762/c  
; Sequence 1762, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1762  
; LENGTH: 453  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(453)  
US-10-156-761-1762

Query Match 86.3%; Score 16.4; DB 15; Length 453;  
Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18  
|||||  
Db 372 GTCGACGCCGCCGATGAC 355

RESULT 14  
US-10-425-114-15573/c  
; Sequence 15573, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 15573  
; LENGTH: 646  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3060-027-C5\_FLI  
US-10-425-114-15573

Query Match 86.3%; Score 16.4; DB 13; Length 646;  
Best Local Similarity 94.4%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGACC 19  
|||||  
Db 296 TCGACGCCGCCGATGACC 279

RESULT 15  
US-10-424-599-23643/c  
; Sequence 23643, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:



APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 30330  
LENGTH: 869  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: UC-ZMFLB73010B10\_FLI  
US-10-425-114-30330

Query Match 86.3%; Score 16.4; DB 13; Length 869;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19  
DB 364 TCGACGCCGCCGATGACC 347

RESULT 18  
US-10-425-114-16256/c  
Sequence 16256, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 16256  
LENGTH: 880  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3062-022-E9\_FLI  
US-10-425-114-16256

Query Match 86.3%; Score 16.4; DB 13; Length 880;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19  
DB 422 TCGACGCCGCCGATGACC 405

RESULT 19  
US-10-425-114-1435/c  
Sequence 1435, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 23643  
LENGTH: 702  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)-(702)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_121351C.1  
US-10-424-599-23643

Query Match 86.3%; Score 16.4; DB 13; Length 702;  
Best Local Similarity 94.4%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19  
DB 176 TCGACGCCGCCGATGACC 159

RESULT 16  
US-10-425-114-2101/c  
Sequence 2101, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 2101  
LENGTH: 836  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700203248\_FLI  
US-10-425-114-2101

Query Match 86.3%; Score 16.4; DB 13; Length 836;  
Best Local Similarity 94.4%; Pred. No. 1.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19  
DB 343 TCGACGCCGCCGATGACC 326

RESULT 17  
US-10-425-114-30330/c  
Sequence 30330, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.

```
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1435
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700152962_FLI
US-10-425-114-1435

Query Match      86.3%; Score 16.4; DB 13; Length 882;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      388 TCGACGCCGCCGATGACC 371

RESULT 20
US-10-425-114-12308/c
; Sequence 12308, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 12308
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701166706_FLI
US-10-425-114-12308

Query Match      86.3%; Score 16.4; DB 13; Length 882;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      336 TCGACGCCGCCGATGACC 319

RESULT 21
US-10-425-114-17287/c
; Sequence 17287, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 17287
```

```
; LENGTH: 889
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3069-003-H4_FLI
US-10-425-114-17287

Query Match      86.3%; Score 16.4; DB 13; Length 889;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      396 TCGACGCCGCCGATGACC 379

RESULT 22
US-10-425-114-7159/c
; Sequence 7159, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 7159
; LENGTH: 893
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700620440_FLI
US-10-425-114-7159

Query Match      86.3%; Score 16.4; DB 13; Length 893;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TCGACGCCGCCGATGACC 19
Db      395 TCGACGCCGCCGATGACC 378

RESULT 23
US-10-424-599-120250
; Sequence 120250, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 120250
; LENGTH: 910
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_79594C.1
US-10-424-599-120250
```



Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GTGACGCGCCGATGACC 18  
Db 209 GTGACGCGCCGATGACC 192

## RESULT 28

US-10-369-493-31945/c  
; Sequence 31945, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 31945  
; LENGTH: 1263  
; TYPE: DNA  
; ORGANISM: Thermobifida fusca  
US-10-369-493-31945

Query Match 86.3%; Score 16.4; DB 16; Length 1263;

Best Local Similarity 94.4%; Pred. No. 1.3e-02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 TCGACGCGCCGATGACC 19  
Db 236 TCGATCGCCGCGATGACC 219

## RESULT 29

US-10-425-114-1473/c  
; Sequence 1473, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 1473  
; LENGTH: 1309  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700154547\_FLI  
US-10-425-114-1473

Query Match 86.3%; Score 16.4; DB 13; Length 1309;

Best Local Similarity 94.4%; Pred. No. 1.3e-02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 TCGACGCGCCGATGACC 19  
Db 913 TCGACGCGCCGATGACC 896

## RESULT 30

US-10-425-114-2580/c  
; Sequence 2580, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 2580  
; LENGTH: 1470  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700217894\_FLI  
US-10-425-114-2580

Query Match 86.3%; Score 16.4; DB 13; Length 1470;

Best Local Similarity 94.4%; Pred. No. 1.2e-02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 TCGACGCGCCGATGACC 19  
Db 997 TCGACGCGCCGATGACC 980

Search completed: June 20, 2004, 17:46:28

Job time : 122.466 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 515.915 Seconds  
(without alignments)  
1596.226 Million cell updates/sec

Title: US-10-624-714-16

Perfect score: 19

Sequence: 1 gtcgacgcgcgcgatgacc 19

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_btg.\*

3: gb\_in.\*

4: gb\_ov.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_in.\*

18: em\_mu.\*

19: em\_mu.\*

20: em\_mu.\*

21: em\_mu.\*

22: em\_mu.\*

23: em\_mu.\*

24: em\_mu.\*

25: em\_mu.\*

26: em\_mu.\*

27: em\_mu.\*

28: em\_mu.\*

29: em\_mu.\*

30: em\_mu.\*

31: em\_mu.\*

32: em\_mu.\*

33: em\_mu.\*

34: em\_mu.\*

35: em\_mu.\*

36: em\_mu.\*

37: em\_mu.\*

38: em\_mu.\*

39: em\_mu.\*

40: em\_mu.\*

41: em\_mu.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	19	100.0	401	6	AR149059	Sequence
2	19	100.0	401	6	AR340431	Sequence
C 3	19	100.0	18121	1	AE007182	Mycobacte
C 4	19	100.0	37164	1	MSGY126	Mycobacte
C 5	19	100.0	244800	1	BX842584	Mycobacte
C 6	19	100.0	278492	1	BX248347	Mycobacte
7	17.4	91.6	2534	8	AK072452	Oryza sat
C 8	17.4	91.6	5966	1	SAWY18862	Streptomy
9	17.4	91.6	60687	8	BX842638	Neurospor
C 10	17.4	91.6	139614	8	AC068950	Oryza sat
11	17.4	91.6	154288	2	AP005494	Oryza sat
C 12	17.4	91.6	230138	14	AF232689	Rat cytome
C 13	17.4	91.6	299175	1	AP005023	Streptomy
14	17.4	91.6	299850	1	AP005949	Bradyrhiz
15	17.4	91.6	301666	8	AE017108	Oryza sat
16	17.4	91.6	343473	1	BX640451	Bordetell
17	17.4	91.6	346259	1	BX640435	Bordetell
C 18	17.4	91.6	348171	1	BX640412	Bordetell
C 19	17	89.5	6736	14	ADP279815	Ascovirus
20	17	89.5	11252	1	AE005103	Halobacte
C 21	17	89.5	113170	8	AP003804	Oryza sat
C 22	17	89.5	347950	1	AP003013	Mesorhizo
C 23	16.4	86.3	227	8	AY023290	Oryza sat
C 24	16.4	86.3	810	1	AFAAZU	A.denitrifi
C 25	16.4	86.3	963	6	AX653785	Sequence
C 26	16.4	86.3	1077	6	AX660080	Sequence
C 27	16.4	86.3	1339	5	AB025282	Callus ga
C 28	16.4	86.3	1389	8	AK104317	Oryza sat
C 29	16.4	86.3	1502	8	AK071251	Oryza sat
C 30	16.4	86.3	1560	6	AX366999	Sequence
C 31	16.4	86.3	1796	8	AK065989	Oryza sat
C 32	16.4	86.3	2000	6	AX509451	Sequence
33	16.4	86.3	2105	8	AK071946	Oryza sat
34	16.4	86.3	2217	8	AK072413	Oryza sat
35	16.4	86.3	2431	1	AF099189	Caulobact
36	16.4	86.3	3249	8	AK110103	Oryza sat
C 37	16.4	86.3	5076	5	AB025284	Gallus ga
C 38	16.4	86.3	10352	1	AF546152	Streptomy
C 39	16.4	86.3	10784	1	AE014719	Bifidobac
40	16.4	86.3	10927	1	AE002535	Neisseria
41	16.4	86.3	11122	1	AE006056	Pasteurel
42	16.4	86.3	11649	1	AE001880	Deinococc
C 43	16.4	86.3	12348	1	AE005856	Caulobact
C 44	16.4	86.3	12891	1	AE005033	Mycobacte
C 45	16.4	86.3	15461	1	AE006933	Mycobacte
C 46	16.4	86.3	22511	1	AF254925	Streptomy
C 47	16.4	86.3	32394	3	AC022473	Leishmani
C 48	16.4	86.3	36202	1	AB103463	Streptomy
C 49	16.4	86.3	38585	3	AC011913	Leishmani
C 50	16.4	86.3	39428	1	AF322256	Streptomy
51	16.4	86.3	39730	1	MTCY8D5	Mycobacteri
C 52	16.4	86.3	40570	3	AC073566	Leishmani
C 53	16.4	86.3	40806	1	MSGY140	Mycobacte
C 54	16.4	86.3	71823	2	AC087724	Arabidops
C 55	16.4	86.3	85690	8	AB028618	Arabidops
C 56	16.4	86.3	96100	8	AP004670	Oryza sat
C 57	16.4	86.3	99761	8	AC006161	Arabidops
58	16.4	86.3	110000	2	LMFLCHR36_34	Continuatio
C 59	16.4	86.3	110352	8	AC006220	Arabidops
60	16.4	86.3	113346	8	AC126222	Oryza sat
61	16.4	86.3	118300	8	AP004165	Oryza sat
C 62	16.4	86.3	119798	8	CNS071P6	Oryza sat
C 63	16.4	86.3	120071	8	CNS08CB3	Oryza sat
64	16.4	86.3	131278	2	AC108753	Oryza sat
65	16.4	86.3	141050	8	AC093093	Oryza sat



C 212	15.8	83.2	1762	3	AY227001	Anopheles
C 213	15.8	83.2	1762	8	AK067916	Oryza sat
C 214	15.8	83.2	1769	1	STMSGMAA	Streptomyces
C 215	15.8	83.2	1797	6	AX654663	Sequence
C 216	15.8	83.2	1800	1	SCO251005	Streptomyces
C 217	15.8	83.2	1845	8	AK108309	Oryza sat
C 218	15.8	83.2	1891	8	TKAT92	T.aestivum
C 219	15.8	83.2	1909	8	AK070412	Oryza sat
C 220	15.8	83.2	1929	6	E05034	CDNA encodi
C 221	15.8	83.2	1929	8	RICATPB	Oryza sativ
C 222	15.8	83.2	1964	1	AF237388	Azotobact
C 223	15.8	83.2	1973	8	AK061681	Oryza sat
C 224	15.8	83.2	1981	8	AK100872	Oryza sat
C 225	15.8	83.2	2000	6	AX655619	Sequence
C 226	15.8	83.2	2036	8	AK100857	Oryza sat
C 227	15.8	83.2	2044	8	HV2172309	Oryza sativ
C 228	15.8	83.2	2062	8	AK100027	Oryza sativ
C 229	15.8	83.2	2063	8	AK069323	Oryza sat
C 230	15.8	83.2	2063	6	AX653988	Sequence
C 231	15.8	83.2	2118	6	AX653988	Sequence
C 232	15.8	83.2	2206	8	AK100683	Oryza sat
C 233	15.8	83.2	2265	8	AK107091	Oryza sat
C 234	15.8	83.2	2271	1	AF217543	Sinorhizo
C 235	15.8	83.2	2372	8	AK106832	Oryza sat
C 236	15.8	83.2	2387	1	RSLGTDGEN	R.sphaeroid
C 237	15.8	83.2	2387	8	AK067392	Oryza sat
C 238	15.8	83.2	2471	1	STHBTYBO	S.thermophi
C 239	15.8	83.2	2471	6	E04368	gDNA encodi
C 240	15.8	83.2	2480	8	AK194939	Oryza sat
C 241	15.8	83.2	2554	1	BR303064	Bordetelli
C 242	15.8	83.2	2554	1	BR303065	Bordetelli
C 243	15.8	83.2	2554	1	BR303066	Bordetelli
C 244	15.8	83.2	2610	1	STVAHPCFA	Salmonella
C 245	15.8	83.2	2633	8	AK072982	Oryza sat
C 246	15.8	83.2	2841	8	AK072204	Oryza sat
C 247	15.8	83.2	2774	1	AF087458	Arcylobac
C 248	15.8	83.2	2866	8	AK072070	Oryza sat
C 249	15.8	83.2	2967	6	AX653888	Sequence
C 250	15.8	83.2	3148	1	SNPCA82	Sinorhizo
C 251	15.8	83.2	3248	8	AK071978	Oryza sat
C 252	15.8	83.2	3302	8	AF509747	Oryza sat
C 253	15.8	83.2	3406	8	AK072337	Oryza sat
C 254	15.8	83.2	4003	1	AB010087	Pseudomon
C 255	15.8	83.2	4426	6	AX773998	Sequence
C 256	15.8	83.2	4500	8	AX461703	Chlamydom
C 257	15.8	83.2	4585	8	AF509749	Oryza sat
C 258	15.8	83.2	4613	8	AF509755	Oryza sat
C 259	15.8	83.2	4615	8	AF509756	Oryza sat
C 260	15.8	83.2	4615	8	AF509759	Oryza sat
C 261	15.8	83.2	4618	8	AF509750	Oryza sat
C 262	15.8	83.2	4621	8	AF509757	Oryza sat
C 263	15.8	83.2	4622	8	AF509753	Oryza sat
C 264	15.8	83.2	4622	8	AF509758	Oryza sat
C 265	15.8	83.2	4626	8	AF509751	Oryza sat
C 266	15.8	83.2	4627	8	AF509752	Oryza sat
C 267	15.8	83.2	4632	8	AF509760	Oryza sat
C 268	15.8	83.2	4637	8	AF509754	Oryza sat
C 269	15.8	83.2	4649	6	AX362790	Sequence
C 270	15.8	83.2	4752	1	MTU38939	Mycobacteri
C 271	15.8	83.2	4857	1	AF143948	Pseudomon
C 272	15.8	83.2	5193	1	MSU050335	Mycobacteri
C 273	15.8	83.				

C 285	15.8	83.2	7784	1	RTR291603	AJ291603	Rhizobium
C 286	15.8	83.2	8009	1	RCATP10FN	X99599	R. capsulatu
C 287	15.8	83.2	8009	6	BD129570	BD129570	Polynucle
C 288	15.8	83.2	8617	1	AE006023	AE006023	Caulobact
C 289	15.8	83.2	8808	1	AE005814	AE005814	Caulobact
C 290	15.8	83.2	9531	1	AY033994	AY033994	Streptomy
C 291	15.8	83.2	10173	1	AE005054	AE005054	Halobacte
C 292	15.8	83.2	10254	1	AE005942	AE005942	Caulobact
C 293	15.8	83.2	10393	1	AE005136	AE005136	Halobacte
C 294	15.8	83.2	10393	1	AE011957	AE011957	Xanthomon
C 295	15.8	83.2	10440	1	AE005714	AE005714	Caulobact
C 296	15.8	83.2	10470	1	AE009163	AE009163	Agrobacte
C 297	15.8	83.2	10486	1	AE008129	AE008129	Agrobacte
C 298	15.8	83.2	10514	1	AF546155	AF546155	Streptomy
C 299	15.8	83.2	10526	1	AE001921	AE001921	Deinococc
C 300	15.8	83.2	10620	1	AE005154	AE005154	Halobacte

## ALIGNMENTS

RESULT 1  
 ARI149059  
 LOCUS  
 DEFINITION  
 Sequence 75 from patent US 6228371.  
 ARI149059  
 ACCESSION  
 VERSION  
 ARI149059.1 GI:15113850  
 KEYWORDS  
 SOURCE  
 Unknown.  
 ORGANISM  
 Unknown.  
 UNCLASSIFIED.  
 REFERENCE  
 1 (bases 1 to 401)  
 AUTHORS  
 Nano.F.E.  
 TITLE  
 Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides  
 JOURNAL  
 Patent: US 6228371-A 75 08-MAY-2001;  
 FEATURES  
 Location/Qualifiers  
 1..401  
 /organism="unknown"  
 /mol\_type="unassigned DNA"  
 source  
 ORIGIN

```

Query Match      100.0%; Score 19; DB 6; Length 401;
Best Local Similarity 100.0%; Pred. NO. 1.2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GTCCAGCGCCGCCGATGACC 19
Db      12  GTCCAGCGCCGCCGATGACC 30

```

RESULT 2					PAT 17-AUG-2003
AR340431					
LOCUS	AR340431	401 bp	DNA	linear	
DEFINITION	Sequence 75 from patent US 6572865.				
ACCESSION	AF340431				
VERSION	AR340431.1	GI:33731979			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	Nano,F.E.				
TITLE	Mycobacterium tuberculosis DNA sequences encoding immunostimulatory peptides and methods for using same				
JOURNAL	Patent: US 6572865-A 75 03-JUN-2003;				
FEATURES	Location/Qualifiers				
source	1..401				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN					
Query Match	100.0%;	Score 19;	DB 6;	Length 401;	

Best Local Similarity 100.0%; Pred. No. 1.2e+03; Mismatches 0; Indels 0; Gaps 0;  
Matches 19; Conservative 0

QY 1 GTCGAGCGCGCGATGACC 19  
|||||  
Db 12 GTCGAGCGCGCGATGACC 30  
|||||

RESULT 3  
AE007182/c  
LOCUS  
DEFINITION  
Mycobacterium tuberculosis CDC1551, section 268 of 280 of the complete genome.  
ACCESSION  
AE007182 AE000516  
VERSION  
AE007182.1 GI:13883751  
KEYWORDS  
SOURCE  
ORGANISM  
Mycobacterium tuberculosis CDC1551  
Mycobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.  
REFERENCE  
1 (bases 1 to 18121)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
Unpublished  
2 (bases 1 to 18121)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Direct Submission  
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
LOCATION/Qualifiers  
1. :18121  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
207. .896  
/gene="WT3873"  
207. .896  
/gene="WT3873"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48237.1"  
/db\_xref="GI:13883752"  
/translation="VRSFDSGLTFGIVTYARNWNANVTYRSMIDAAGGLHPV  
ALMLVEGGNPPGSGSWINLYNLYADYAGSPVLLIGYANVDFNVEYRPAGLR  
VIGAGYGNFNPFGVQVHQYTDGSGYSPNLFQAGPPFCRDMNSANGLTPOQFAAAG  
VTTCGGPLNALDEEQTETLLTKAREINDQLRPGNAGWPQUGNEQGDUTFPVDAIAV  
IKNDVAAMLAIE"  
complement(910. .1854)  
/gene="WT3874"  
complement(910. .1854)  
/gene="WT3874"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK48238.1"  
/db\_xref="GI:13883753"  
/translation="WPRDNDNSWALTESVGTALGVAAAPAAETESNPLNDPFARI  
FYDAAGDGIWSMYNRTLLAGATDLDPLRAPICQMIDFMAARTAFDEYFLATADAG

gene  
CDS  
1884. .2343  
/gene="WT3875"  
1884. .2343  
/gene="WT3875"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48239.1"  
/db\_xref="GI:13883754"  
/translation="MGSTPPTPQEVFAHHGOALAGDLDEIVADYADDSFVITPAGI  
ARGKEGIROLFVKLLDDIPNALWDLKTIQIFEGDILFLEWTANSASVSRVDDGVDVTFVR  
DGTIAHTVRYTPHPTK"  
complement(2388. .2906)  
/gene="WT3876"  
complement(2388. .2906)  
/gene="WT3876"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48240.1"  
/db\_xref="GI:13883755"  
/translation="MTPSGTTRNHGVSRIHRPQITVRCRAQKIAQFPFSGESLVEGABQK  
VLIDGLSLFORPHEPRIGVGLGRELAFGRHVAQVSGVPFDCRRDFSLVGRARVDGR  
DGSIVGLVRFQSGDPCSLLQRHGLILSGRSAGGSACKGCNLTSCAARNAPPGGI  
RRQITVPRSDA"  
2996. .3127  
/gene="WT3876.1"  
2996. .3127  
/gene="WT3876.1"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48241.1"  
/db\_xref="GI:13883756"  
/translation="MNRTWGCADIGWRYGLEVADPVRARGRHPSPDVEDRALITWR"  
complement(3114. .3689)  
/gene="WT3877"  
complement(3114. .3689)  
/gene="WT3877"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48242.1"  
/db\_xref="GI:13883757"  
/translation="MLSGIQNTLMDNDPLANGYVADLLVALAVLMLRARRTERPE  
LAPMLLGLTGLVWELPVFGSLAWNTLPIEWAIPPLPFTVTVFLLAHSWDGPLITM  
GMLRALTLGEPAGALGLTVQVWLQLTALAVELSAIILAGTWSYVDDLNFNPFMEWFR  
GHPVTAAMGLTWLLAPLCLPFAALVRLALTAR"  
complement(3795. .3977)  
/gene="WT3878"  
complement(3795. .3977)  
/gene="WT3878"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48243.1"  
/db\_xref="GI:13883758"  
/translation="MGSTPWCNQCQTLRTPTVEVLELAVALRPNPDRTAGAIQRIL  
RAQAGDRIALRGSGS"  
complement(3992. .4183)  
/gene="WT3879"  
complement(3992. .4183)  
/gene="WT3879"  
/note="identified by Glimmer2; putative"

gene  
CDS  
1884. .2343  
/gene="WT3875"  
1884. .2343  
/gene="WT3875"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48243.1"  
/db\_xref="GI:13883758"  
/translation="MGSTPWCNQCQTLRTPTVEVLELAVALRPNPDRTAGAIQRIL  
RAQAGDRIALRGSGS"  
complement(3992. .4183)  
/gene="WT3879"  
complement(3992. .4183)  
/gene="WT3879"  
/note="identified by Glimmer2; putative"

gene  
CDS  
1884. .2343  
/gene="WT3875"  
1884. .2343  
/gene="WT3875"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK48243.1"  
/db\_xref="GI:13883758"  
/translation="MGSTPWCNQCQTLRTPTVEVLELAVALRPNPDRTAGAIQRIL  
RAQAGDRIALRGSGS"  
complement(3992. .4183)  
/gene="WT3879"  
complement(3992. .4183)  
/gene="WT3879"  
/note="identified by Glimmer2; putative"



/codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAK48244.1"  
 /db\_xref="GI:13883759"  
 /translation="MRAERARA GLFRYQLIREADAHAHSTKRGKQVRELASREHTD  
 PFGKVRISRTIDRWIRN"  
 complement (4318. .4644)  
 /gene="WT3880"  
 complement (4318. .4644)  
 /gene="WT3880"  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAK48245.1"  
 /db\_xref="GI:13883760"  
 /translation="MPAPAKALSOVGFRIADLPAPETVGRGLERFAERAEAVRS  
 VFTVLRVADPDPMPDRAVGVFAVTVIAAVVTVIECFALSTVSLATAVAVSGG  
 RLVAAG"  
 complement (4779. .4851)  
 /gene="tRNA-Arg-4"  
 complement (4779. .4851)  
 /gene="tRNA-Arg-4"  
 /product="tRNA-Arg"  
 complement (4882. .4970)  
 /gene="tRNA-Ser-3"  
 complement (4882. .4970)  
 /gene="tRNA-Ser-3"  
 /product="tRNA-Ser"  
 5048. .6109  
 /gene="WT3881"  
 5048. .6109  
 /gene="WT3881"  
 /note="similar to GP:5712092; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="histidinol-phosphate aminotransferase"  
 /protein\_id="AAK48246.1"  
 /db\_xref="GI:13883761"  
 /translation="MTARLRPELAGLPVYVPGKTPVGAIKLASNETVFGPLPSVRAAI  
 DRATVNVYDNGCVOLKAALRHLPDPFAPEHVAVCGSVSLCQOLVQVTVASVGE  
 TVFGSRPELPVQVAVAGAPIQVPLTDHFDLYAVLATVDRTRLI FVCPNPNPTS  
 TVVGDALAREVAPAHILIAIDEAVVEVIRDMRPDPSGLVRAHNVVVLRTFKA  
 YGLAGLRIGYAGHPDVIATLDRKIVVFTVSSIGCAAAIASLDADELLARTDTVAE  
 RARVSAELRAAGFTLPSPQANFVWLPLGSRQDFEQADARIIVRPYGTGVRVTVA  
 APEADAPLRPARWRSDQ"  
 complement (6155. .6739)  
 /gene="WT3882"  
 complement (6155. .6739)  
 /gene="WT3882"  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAK48247.1"  
 /db\_xref="GI:13883762"  
 /translation="MPESRGRDPSPTDELACAAALQVLTHTIGRQDKAKQTP  
 CPGDVKILTHLNLINVLGMVGAFLSRADIDSVRLVSGHSAALDAWRHGLPE  
 GDVSLGPGMSAKVAVSVFSEFLVHAWDYAVAVSELKAADSLAEYVLELARKLKP  
 EERSVAGNEFVDPEDGGALERLIAFTGRNPAR"  
 6763. .7587  
 /gene="WT3883"

Query Match 100.0%; Score 19; DB 1; Length 18121;  
 Best Local Similarity 100.0%; Pred. No. 9.4e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCGACGCCGCGATGACC 19  
 |||||  
 Db 14977 GTCGACGCCGCGATGACC 14959

RESULT 4  
 MSGV126/c 37164 bp DNA linear BCT 10-DEC-1996  
 LOCUS Mycobacterium tuberculosis sequence from clone y126.  
 DEFINITION  
 ACCESSION AD000012  
 VERSION AD000012.1 GI:1717733  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mycobacterium tuberculosis  
 Mycobacterium tuberculosis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
 1 (bases 1 to 37164)  
 Du, L.  
 Direct Submission  
 TITLE Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100  
 Reaver Street, Waltham, MA, USA, 02154 du@ctic.com  
 COMMENT GSDB:S:1004704  
 FEATURES  
 Location/Qualifiers  
 1..37164  
 /organism="Mycobacterium tuberculosis"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1773"  
 /clone="y126.nc"  
 ORIGIN  
 Query Match 100.0%; Score 19; DB 1; Length 37164;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCGACGCCGCGATGACC 19  
 |||||  
 Db 13712 GTCGACGCCGCGATGACC 13694  
 RESULT 5  
 BX842584/c 244800 bp DNA circular BCT 21-NOV-2003  
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 13/13  
 DEFINITION  
 ACCESSION BX842584 AL022076 AL022076 AL022120 AL023456 Z80343 Z83864  
 Z94121 Z97188  
 VERSION BX842584.1 GI:38490370  
 KEYWORDS complete genome.  
 SOURCE Mycobacterium tuberculosis H37Rv  
 ORGANISM Mycobacterium tuberculosis H37Rv  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium  
 tuberculosis complex.  
 1  
 Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,  
 Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,  
 Tekaita, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,  
 Connor, R., Davies, K., Devlin, K., Feltwell, T., Gentles, S.,  
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,  
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,  
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,  
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and  
 Barrall, B.G.  
 TITLE Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence  
 JOURNAL Nature 393 (6685), 537-544 (1998)  
 MEDLINE 98295987  
 PUBMED 9634230  
 REFERENCE  
 2  
 Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.  
 Re-annotation of the genome sequence of Mycobacterium tuberculosis  
 H37Rv  
 JOURNAL Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)  
 MEDLINE 22255591  
 PUBMED 12368430  
 REFERENCE  
 3 (bases 1 to 244800)  
 Parkhill, J.

TITLE JOURNAL	Direct Submission Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk On or before Nov 21, 2003 this sequence version replaced gi:3261511, gi:3256026, gi:3261558, gi:3261559, gi:3261648, gi:3261687, gi:3261736, gi:3261805.
COMMENT	Notes: Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, <a href="http://www.sanger.ac.uk/Projects/M_tuberculosis/">http://www.sanger.ac.uk/Projects/M_tuberculosis/</a> ).
FEATURES source	1. .244800 /organism="Mycobacterium tuberculosis H37Rv" /mol_type="genomic DNA" /strain="H37Rv" /db_xref="taxon:83332" complement(89. .1396) /locus_tag="RV3722c" complement(89. .1396) /locus_tag="RV3722c" /function="UNKNOWN" /note="RV3722c, (MTV025.070c), len: 435 aa. Conserved hypothetical protein, equivalent to O69513 MLCB2407.14 (alias O9CB73 ML2336, 463 aa) HYPOTHETICAL 46.8 KDA PROTEIN from Mycobacterium leprae (426 aa), FASTA scores: opt: 2505, E(): 8.3e-154, (87.25% identity in 424 aa overlap). Also highly similar to O9RUI7 DR1579 CONSERVED HYPOTHETICAL PROTEIN from Deinococcus radiodurans (452 aa), FASTA scores: opt: 1162, E(): 3.1e-67, (44.8% identity in 422 aa overlap); and partially similar to O9I371 PA1654 PROBABLE AMINOTRANSFERASE from Pseudomonas aeruginosa (388 aa) FASTA scores: opt: 162, E(): 0.0078, (25.85% identity in 348 aa overlap) and other aminotransferases. Ttparse score is 0.900. N-terminus extended since first submission (previously 408 aa)." /codon_start=1 /evidence="experimental" /transl_table=11 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAA18044.2" /db_xref="GI:38490371"
gene	/product="PROBABLE CONSERVED TRANSMEMBRANE PROTEIN" /protein_id="CAA18045.1" /db_xref="GI:2960147" /translation="MGRKVALVLMHSAFSGAGLYFYFVLPWPEIMGDTGHSGLTGLRIATGALVGLAALPVVFTLLRTRKPELGPQALSMRIWSINAHVLAGALIVGTISGLVMSLDAAGQWLFGLYGAATAVAVGFGFVLSFVAELPPPKPKPKPKQRLRRKXTAKGDEAPRAEAETELAAQDEAEVAPPEIESPGEGESATREAPAETA TASEPKGRNRRPTGKTSRRRRIRKSGVQVAKVDE"
CDS	2735. .2977 /genes="cut5a" /locus_tag="RV3724A" 2735. .2977 /genes="cut5a" /locus_tag="RV3724A" /EC_number="3.1.1.1." /function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)." /note="RV3724A, (MTV025.072), len: 80 aa. Probable cut5a, truncated cutinase precursor (EC 3.1.1.-), similar to N-terminal end of others e.g. O9KK87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 202, E(): 1.5e-06, (56.45% identity in 52 aa overlap); O9XB09 RV22-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa), FASTA scores: opt: 200, E(): 1.5e-06, (61.4% identity in 57 aa overlap); and O00298 CUTI_BOTCI CUTA CUTINASE PRECURSOR from Botrytis cinerea (Botryotinia fuckeliana) (202 aa), FASTA scores: opt: 108, E(): 2.2, (40.4% identity in 52 aa overlap). Also highly similar to others from Mycobacterium tuberculosis e.g. O06318 CUT3 MYCTU RV3451 MT3557 MTY13E12.04 PROBABLE CUTINASE PRECURSOR (247 aa), FASTA scores: opt: 189, E(): 1.2e-05, (58.0% identity in 50 aa overlap); O50664 CUT2 MYCTU RV2301 MT2358 MTY339.08c PROBABLE CUTINASE PRECURSOR (219 aa), FASTA scores: opt: 172, E(): 0.00015, (59.2% identity in 49 aa overlap); O.00015, (59.2% identity in 49 aa overlap); O06793 RV1758 MTY28.24 295890 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29, (57.2% identity in 166 aa overlap); O06319 RV3452 MTY13E12.05, and U00015.11 from Mycobacterium leprae. BELONGS TO THE CUTINASE FAMILY. Rest of cutinase ORF continues as RV3724B CUT5B, frameshifting could occur near position 4169668. Sequence has been checked but no errors found." /codon_start=1 /transl_table=11 /product="PROBABLE CUTINASE PRECURSOR [FIRST PART] CUT5A" /protein_id="CAE55632.1" /db_xref="GI:38490372" /translation="MDVIRNARRLAVACTRAAVTTPGLLSAHVPMVSAEPCPDVEVV FARGTEPGGIGSVGLFVDALRFFGWRQVTRGLRR"
gene	2874. .3437 /genes="cut5b" /locus_tag="RV3724B" 2874. .3437 /genes="cut5b" /locus_tag="RV3724B" /EC_number="3.1.1.1." /function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)." /note="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b, truncated cutinase (EC 3.1.1.-), similar to C-terminal end of others e.g. O9XB09 RV22-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa) FASTA scores: opt: 335, E(): 3.4e-12, (53.25% identity in 92 aa overlap); O9KK87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05% identity in 168 aa overlap). Also similar to proteins from Mycobacterium tuberculosis e.g. O06793 RV1758 MTY28.24 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29, (57.2% identity in 166 aa overlap). O06319 RV3452 MTY13E12.05 HYPOTHETICAL 23.1 KDA PROTEIN (226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65%
CDS	2874. .3437 /genes="cut5b" /locus_tag="RV3724B" 2874. .3437 /genes="cut5b" /locus_tag="RV3724B" /EC_number="3.1.1.1." /function="HYDROLYSIS OF CUTIN (A POLYESTER THAT FORMS THE STRUCTURE OF PLANT CUTICLE)." /note="RV3724B, (MTV025.072), len: 187 aa. Probable cut5b, truncated cutinase (EC 3.1.1.-), similar to C-terminal end of others e.g. O9XB09 RV22-RV1758 PROTEIN (FRAGMENT) from Mycobacterium bovis BCG (143 aa) FASTA scores: opt: 335, E(): 3.4e-12, (53.25% identity in 92 aa overlap); O9KK87 SERINE ESTERASE CUTINASE from Mycobacterium avium (220 aa), FASTA scores: opt: 251, E(): 2.5e-07, (44.05% identity in 168 aa overlap). Also similar to proteins from Mycobacterium tuberculosis e.g. O06793 RV1758 MTY28.24 HYPOTHETICAL 17.9 KDA PROTEIN (174 aa), FASTA scores: opt: 641, E(): 2.7e-29, (57.2% identity in 166 aa overlap). O06319 RV3452 MTY13E12.05 HYPOTHETICAL 23.1 KDA PROTEIN (226 aa), FASTA scores: opt: 385, E(): 7.5e-15, (46.65%
gene	1613. .1698 /genes="CRNA-Ser (GGA) " 1613. .1698 /genes="CRNA-Ser (GGA) " /product="CRNA-Ser" /note="codon recognized: UCC" /anticodon=" (pos:1647. .1649,aa:Ser) 1804. .2568 /locus_tag="RV3723" 1804. .2568 /locus_tag="RV3723" /function="UNKNOWN" /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Ttparse score is 0.900." /codon_start=1 /transl_table=11
gene	1613. .1698 /genes="CRNA-Ser (GGA) " 1613. .1698 /genes="CRNA-Ser (GGA) " /product="CRNA-Ser" /note="codon recognized: UCC" /anticodon=" (pos:1647. .1649,aa:Ser) 1804. .2568 /locus_tag="RV3723" 1804. .2568 /locus_tag="RV3723" /function="UNKNOWN" /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Ttparse score is 0.900." /codon_start=1 /transl_table=11
tRNA	1613. .1698 /genes="CRNA-Ser (GGA) " 1613. .1698 /genes="CRNA-Ser (GGA) " /product="CRNA-Ser" /note="codon recognized: UCC" /anticodon=" (pos:1647. .1649,aa:Ser) 1804. .2568 /locus_tag="RV3723" 1804. .2568 /locus_tag="RV3723" /function="UNKNOWN" /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Ttparse score is 0.900." /codon_start=1 /transl_table=11
gene	1613. .1698 /genes="CRNA-Ser (GGA) " 1613. .1698 /genes="CRNA-Ser (GGA) " /product="CRNA-Ser" /note="codon recognized: UCC" /anticodon=" (pos:1647. .1649,aa:Ser) 1804. .2568 /locus_tag="RV3723" 1804. .2568 /locus_tag="RV3723" /function="UNKNOWN" /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Ttparse score is 0.900." /codon_start=1 /transl_table=11
CDS	1613. .1698 /genes="CRNA-Ser (GGA) " 1613. .1698 /genes="CRNA-Ser (GGA) " /product="CRNA-Ser" /note="codon recognized: UCC" /anticodon=" (pos:1647. .1649,aa:Ser) 1804. .2568 /locus_tag="RV3723" 1804. .2568 /locus_tag="RV3723" /function="UNKNOWN" /note="RV3723, (MTV025.071), len: 254 aa. Probable conserved transmembrane protein, with hydrophobic stretches at the N-terminus, and equivalent to O69512 MLCB2407.13c PUTATIVE MEMBRANE PROTEIN from Mycobacterium leprae (250 aa), FASTA scores: opt: 1029, E(): 1.2e-44, (64.45% identity in 253 aa overlap). Ttparse score is 0.900." /codon_start=1 /transl_table=11





Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.

Direct Submission  
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,  
Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT  
This clone is one of the 28k full-length cDNA clones from japonica rice.  
URL : <http://cdna01.dna.affrc.go.jp/cDNA/>  
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and Yamamoto,M.

FAIS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kuroaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Nariakawa,R., Niikura,J., Oka,M., Ryu,K., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kouda,M., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

Location/Qualifiers  
1..2554  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultiivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J023121D13"

ORIGIN  
Query Match 91.6%; Score 17.4; DB 8; Length 2554;  
Best Local Similarity 94.7%; Pred. No. 4.7e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTCGACGCGCGGATGACC 19  
|||||  
DB 2256 GTCGACGCGCGGATGACC 2274  
|||||

RESULT 8  
SAMV18862/c  
LOCUS SAMV18862 5966 bp DNA linear BCT 03-FEB-2000  
DEFINITION Streptomyces ambofaciens NRRL2240 sar gene and ORF2, ORF3 and ORF4.  
ACCESSION Y18862  
VERSION Y18862.1 GI:6911695  
KEYWORDS polyketide synthase; sar gene; Streptomyces ambofaciens regulator; thioesterase  
SOURCE Streptomyces ambofaciens  
ORGANISM Streptomyces ambofaciens  
Bacteria; Actinobacteridae; Actinobacteriales;  
Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Culebras,E., Martinez,E., Carnero,A. and Malpartida,F.

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
JOURNAL  
TITLE  
JOURNAL  
FEATURES  
source  
gene  
CDS  
CDS  
CDS  
CDS

Cloning and characterization of a regulatory gene of the SARP family and its flanking region from Streptomyces ambofaciens  
Mol. Gen. Genet. 262 (4-5), 730-737 (1999)  
20092477  
10628855  
2 (bases 1 to 5966)  
Malpartida,F.  
Direct Submission  
Submitted (11-MAR-1999) F. Malpartida, Centro Nacional de Biotecnología, Microbiología, Lab 218, Campus de la UAM 28049 Canto Blanco, Madrid, 28049, SPAIN  
Location/Qualifiers  
1..5966  
/organism="Streptomyces ambofaciens"  
/mol\_type="genomic DNA"  
/strain="NRRL2240"  
/db\_xref="taxon:1889"  
661..1483  
/gene="sar"  
661..1483  
/gene="sar"  
/note="Protein sequence is in conflict with the conceptual translation"  
/codon\_start=2  
/transl\_table=11  
/product="Streptomyces ambofaciens regulator (SAR)"  
/protein\_id="CAB72140.1"  
/db\_xref="GI:6911696"  
/db\_xref="GOA:Q9KWX4"  
/db\_xref="SPTREMBL:Q9KWX4"  
/translation="MKI QVGLNALNAEFNGISVVTAGKPRQILALLALCPGRVVPVT LMEETGTELPQSSMTTQTYTLQRLRLGTAGPVPKSAKDLATRYGVLQIPA EQVADFOYERLVVEGRQAFEDGEDERAALVDAALDMDPAPLVDPVGPPILEIVMR MEQSRIVTERRIDAEILGRHVELIAELTDLVARHPQEGHLSQAMVALYRSQAA ALDVYRRLQRLIELGVPEPQLQRLHQAQLAVDQLDVGAGARRSSTFDLYAA" 1842..2622  
/note="ORF2"  
/codon\_start=2  
/transl\_table=11  
/product="putative thioesterase"  
/protein\_id="CAB72141.1"  
/db\_xref="GI:6911697"  
/db\_xref="GOA:Q9KWX3"  
/db\_xref="SPTREMBL:Q9KWX3"  
/translation="VASRTKAPARPARVTLNCLAHAGAGVASYRDWASAVGVDVH ALPLPGREARRRPRITTEADLLADLLPGLLTARRGPYALYHSGNALVGYTLTRAL ADAGAPPLFVAGCAPPTHTTALVDGADLPDDELLPLDTIGSLPPGASAPGGLWR RTFLPVLRLDLRLAKSLRNAALDPVTGPDVPLVFAAGSDPFLAAPAALRHQQWTT DIELHTVSGGHEFFASSAGLAQHVGACDHVNAQAGRR" 2619..3890  
/note="ORF3"  
/codon\_start=1  
/transl\_table=11  
/product="putative polyketide synthase"  
/protein\_id="CAB72142.1"  
/db\_xref="GI:6911698"  
/db\_xref="GOA:Q9KWX2"  
/db\_xref="SPTREMBL:Q9KWX2"  
/translation="MKRVITIGVVPAGPGVTKSLWDLTLTGRATRPTITFLDAGY RSRIAAEVDFPAAGFAGADTERLDRITQFALVAREAIADSGTGESVERAPLRCAV SLGSAIGCTTALATVAILSDCGNTWDLDTAADFLYDFVPFSSLAALARDCCGAG PVALVSSGCTSLGDAIGHGADLIREGSDMVAUVGGAEPAPITWASFDRLRLTSRN DPATASRPDRTRDPVY.GEGAAVLVLEQSHARRPAGQAPVLAELSLAHSAAHMT GLRPAQEAQAAIRAAALDQARLNADIDYISAHGAGTRNDSHETHAFKAGLGHHAHQ VPVSSIKSMIGHALGAALDITAGALAIRHGIIVPTANLHQPDPFCDLDYTFLFARE QRTSTVLVSAGSGFGSPHTAAVLTRPQLQAA" 3887..5149  
/note="ORF4"  
/codon\_start=1  
/transl\_table=11  
/product="putative polyketide synthase"  
/protein\_id="CAB72143.1"  
/db\_xref="GI:6911699"

```

/db_xref="GOA:Q9KWX1"
/db_xref="SPTREMBL:Q9KWX1"
/translation="MSHPAPPAPSPAPDPGPHPTHQALVTGIGITAPNGLCAPAYWD
ALLSGRGLGPTTRDASGYPVRIAGEITDFATRNTSQOTPASTAAPASPWPRPKPS
TTRGAGPOETAGNAGVVTASLAGAEFGQGLAALWSKGATHSVAYCSFASPHAAAP
AOISIRHLRGHGSAILIIEQAGGIDALARARRRINDGAPLWTVGVDSTLCPWGAAR
LADGRLSPGSEPHHARFFPAADAGHTVGGEGALLVLEDAARSARRGATGTVAGAP
PSTAPRPRNRERALEALAQGLTPAHVVDVADGAGRRAPTAAPRAPHCAPCAPYV
VPTVPTKMTGRGAGGALDVATALLAALNAKTVPITGTTTRVAGDCPLDVTAPRRL
PHLQVALVLAADAEAASTAATVLRAPHG"

ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 5966;
Best Local Similarity 94.7%; Pred. No. 4.4e-03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGAGCGCCGCGATGACC 19
|||||
Db 4507 GTCGAGCGCCGCGATGACC 4489

RESULT 9
BX842638 Neurospora crassa DNA linear PLN 30-NOV-2003
LOCUS Neurospora crassa DNA linkage group I cosmid contig G17B7.
DEFINITION BX842638
ACCESSION BX842638.1 GI:38567282
VERSION
KEYWORDS
SOURCE Neurospora crassa
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezilomycotina; Sordariomycetes;
Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 Schulte, U., Aign, V., Hoheisel, J., Brandt, P., Fartmann, B.,
Holland, R., Nyakatura, G., Mewes, H.W. and Mannhaupt, G.
Unpublished
2 (bases 1 to 60687)
German Neurospora genome project.
Direct Submission
Submitted (27-NOV-2003) MIPS, Institut fuer Bioinformatik,
Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail:
G.Mannhaupt@sf.de Project Coordinator: Ulrich Schulte, Institute
of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,
E-mail: ulrich-schulte@uni-duesseldorf.de
this contig is an assembly of cosmid G17B7 from 1 to: 38101, cosmid
library pMOcosX, strain 74-OR-23-1VA, and cosmid 87E11 from 38102
to: 60687, pLORIS6Xn, strain 74-OR-23-1A, cosmid clones are
available at the Fungal Genetic Stock Center, http://www.fgsc.net
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
http://www.mwgdna.com
Information on performance of analysis and a more detailed
annotation of this entry and other sequences can be viewed at:
http://mips.gsf.de/proj/neurospora.
FEATURES
Location/Qualifiers
1..60687
/organism="Neurospora crassa"
/mol_type="genomic DNA"
/db_xref="taxon:5141"
/chromosome="I"
1657..3815
/gene="G17B7.010"
complement(join(1657..3038,3143..3539,3810..3815))
/gene="G17B7.010"
complement(join(1657..3038,3143..3539,3810..3815))
/gene="G17B7.010"
/note="match of 17 ESTs
similarity to lysyl-tRNA synthetase (EC 6.1.1.6), Homo
sapiens, SWISSPROT:SYK_HUMAN
contains EST gb:AW710587, AW710588"
/codon_start=1
/product="related to lysyl-tRNA synthetase (lysine-tRNA
ligase)"
/protein_id="CAE76572.1"
/db_xref="GI:38567283"
/translation="MIFLRPYAKWKPKVGARNTAYLRFYHARQHHELIQSNALKYPRI
RHEGGAMPITPFRKFKDVENMGKHADESVI IGRVOSVRRAGNKLMTDLKGEFEHV
QGLCNFKLVQTVGVELGEFKALAKLLNRGDIVSLTGRAKTKPTGELSIEATRLPELLT
PALAPLPKLEDEEVRIQHRHIDMLVNRKIDITDLRLRSHLIXYLRDFFHFERGFLSEFQT
PIAENASGATARPFTVTASSETSVNKNLWRIAPLRLWKLRLVGGVGVKVFELGAPFRN
EGDQTHNPFCEFNCFYNAYANLDDILITLIEELICGAASHCRDLIATKLDSPADMS
QYRQPFLEFIPALSAEAGFKLPDLSASSREDLIALKQHKIEVPGLNKPDITLAK
LIDLNLSELEPYSLTPLFIHTHPVCMSPLSKSFVCPKTHQHVARSARTELFNGKELA
NMYEBENDPASOROKFVDOLNAKREASLAPGVYNDEGSAVEVESVVOALESGLPP
TCGWGCGIERMWMLESCTKEISDCLSFGLNRHVTNVPGARVTGEDEEKAREMWEVEK
KGARERKREKKEKEDMGVSDGEA"
complement(1657..3038)
/gene="G17B7.010"
/number=1
complement(3039..3142)
/gene="G17B7.010"
/number=1
complement(3143..3539)
/gene="G17B7.010"
/number=2
complement(3540..3809)
/gene="G17B7.010"
/number=2
complement(3810..3815)
/gene="G17B7.010"
/number=3
4460..4985
/gene="G17B7.020"
join(4460..4551,4805..4985)
/gene="G17B7.020"
/note="identical to TREMBL:AF343075.1
contains EST gb:AW714171, AW709501, AW712102, AW708270,
AW711482, AI329860"
/codon_start=1
/product="mitochondrial import inner membrane translocase
subunit TIM10"
/protein_id="CAE76573.1"
/db_xref="GI:38567284"
/translation="MFGICRPOPTSAEKIAVAENELKVAENHSHRMVICILKCIDKS
YREGDUSKGESVLCRCAAKFFETHQKISDQLQKQETQARGGGGFGM"
4460..4551
/gene="G17B7.020"
/number=1
4552..4804
/gene="G17B7.020"
/number=1
4805..4985
/gene="G17B7.020"
/number=2
5985..9947
/gene="G17B7.030"
complement(5985..9947)
/gene="G17B7.030"
complement(5985..9947)
/gene="G17B7.030"
/note="contains EST gb:BG280456, BG280455"
/codon_start=1
/product="putative protein"
/protein_id="CAE76574.1"
/db_xref="GI:38567285"
/translation="MTDDPASPNSLRPRRTATKQPDLDSDRHDYASTAISRQDHVTT
NLPDVSVTSSDLNDVKDEYDDASIPDQRLPRSHPRADRNFRSGALRAHKHRTNGA
FLISDPALGANTQHTQLQPKRQSLLEHGRNITKNGMDRSPGRNKLTPGNEVLGGL
VORSSTTPSVRQPTSDINSAVTNRRSGSAGETIIGTNPSPRTSMALLDMESAIVN
MALNLSERRRIASRRIVSPVPRPGVPENSTGSLRQHLQOQSKVRNNSPRDHS
PRIASASLQASAGLEAVQYHESQSTLARAQKAREVGLMACQYRELLLPPLKAPS
ATYSVSPPLTNGFNFRPTPTHDAETIGPYNPLOYIENRKIRAREKTLDEGQ
GNDVSEVSEWDEVAKMIAKQDKRIPGNPSIPFSPTRILUSCEIPFQSLGSKPTPT
PKLKKPRVDVAVDPAIDMIADYWLQDDDNKFLVEDRNRWRLFPQGSATYRSLSRDT
POSSKIFNSNSITALKARTQSPQVVEPKTKAHQEPHPGFSASARAKAQQKLR
ATREHSHRANSIYNRDLIRHRRSFSESSDSDSDRKRGRHGTISSSGKDIARQME"

```

DMIAEQORD:EAHPLVDQDQOQLKPLDAGLAMSEKQGWPRITSTSDGADNVEGRPPA  
SRPAPADESNTKYQAPAPQVLAGRAGVLSGDSGRISTDYDTSPPYSDPLRPVH  
DALIPRESTOSNRSRSPPIARSKNELTKVRSITFRDSNEHLAELPSHIGRDN  
PESTQFYLDQWPEPTLLGFGTGMSPKPKTSRPLKRVITVKTGSDSOKNGSTGI  
YKJRDGNSLKGELFGORIDSVLRSGSVKSDMFWRKEAPGESDKYSDSDSDAES  
RGRSRLPRSPKVSRSRPNHEKTKNGEKYLDVMPQFVPTAKHHSKSKSVGSDIDP  
NGLSASPQDQSPRSELLPKPIDIQNASPTSSPGLEITLALRFPWSDVDFDV  
EGKRSKSTVQGVQDADARLNASLMAXGROLVNSASAGWSIAADRENNATPVPP  
QTIVSKREIARLALLSSGISHAMEIDRRAKADKLNASNKTDDEANNTTASSHNN  
NSRLSPDPIVALSPDSTROQLVPIAQIDLFPVAANILSTIQNSATAFQSSVAF  
QTQTAPMTNVRVRLSRISSELTETRRAAEADAGDGLMTHORLKVVRVYDMIDR  
VNRERRERFVRAGWLAWEVLVGFVWFVFMARVLLVGRGVGVGRWLLWL  
complement(5985. .9947)  
/gene="G17B7.030"  
/number=1  
11752. .13517  
/gene="G17B7.040"  
complement(join(11752. .11811,11907. .13517))  
/gene="G17B7.040"  
complement(join(11752. .11811,11907. .13517))  
/gene="G17B7.040"  
/note="similarity to unknown protein AT3g51130,  
Arabidopsis thaliana, TREMBL:AY037194.1"  
/codon\_start=1  
/product="conserved hypothetical protein"  
/protein\_id="CAE76575.1"  
/db\_xref="GI:38567286"  
/translation="MSTPTSPSPFLFSACLYPGRALGVLAGSLHDLTLRLKTEP  
QKFNLDVKSPTPTSIIVIDLIPANGIRLOFQDAPEQLRLIEVVDFTKNHIFIKAA  
DRKEDLVRSASASTPTDNPAGPTFRHIVQRLGPDYDFGFIQDPTDNADVQGL  
VLSWPGVAFSLKMKKSYSAKKNVLDLFPSTQVETMIFSGDSMTQARTQVLDVL  
PYLTKFTPLKGVFDEYSLIKIAGAGKLOVFRKTDQHFWRIGRQTPDVLVSHL  
GPPDAIRKNDHAKFAHFKRPSDALRPSADLKHQDDTLTDQSSANAGSGYSS  
TDESDSVGHVSGECFYFYHGFVDVLVSKTAPSPVFPQEQKNVNPIDRMLL  
TANASIVATKIVLHGVGVSYPFNHRRCKWEIPVLSASDVINSEDFPQESL  
RKQWQSTYDLEKQAGKQVNLNMGWSDSGSIEFLGGWEGDGAAPADAGTAKK  
FDGPADSTLYGPGVLVFEVLKNGAVTALVLRDTLLIGGAGYMWNG"  
complement(11752. .11811)  
/gene="G17B7.040"  
/number=1  
complement(11812. .11906)  
/gene="G17B7.040"  
/number=1  
complement(11907. .13517)  
/gene="G17B7.040"  
/number=2  
14789. .18540  
/gene="G17B7.050"  
join(14789. .15271,15322. .17478,17536. .18540)  
/gene="G17B7.050"  
/note="strong similarity to kinesin-related protein bimC,  
Emicella nidulans, PIR:A34795"  
/codon\_start=1  
/product="probable kinesin-related protein bimC"  
/protein\_id="CAE76576.1"  
/db\_xref="GI:38567287"  
/translation="MERKPSMSRPPSVAGTRSNSSMRPPTRNTASSTGROSALRN  
ASAGGSRABSGAASPADQVSATAGAKERDFDPEDRGTTNNVVRCGRNDRE  
VKNSAVYKTEGKGLVLSMGNVSNKTYTFDFVSCAQADQVDFVVKPIL  
EMLAGNYCTPAYQGTGKTYTWSGDMNQTFGLPDTAGIIFRVLHALNKLVDK  
EHCVRGFIELNBEELDLAEDNTKLIPDNNKGGHATTVQGMERHILSADG  
LWQEGSVRRQVATKNDLSRSHVTFITVHKQOTNGEDYLMGKXNLVLDLAG  
SENIRGAENKAAEAGLNLKSLTIGRVINALVDSPIPYRESKTLRLQDSLGG  
RTKCIATISPASNVEETISLDYAFRAKRNKPOINALLNKTLLREFATEIEK  
LKSLLIATQNGVYLNDAVEELTVONSRRLITEQAAIRIETLVNLRNKQELYT  
MSTFMGLKREHATKQDQDKVLOTVELVLAATKSLAEAHLEHRAHQTOKLA  
KVBGLISTVKTQVDEGLHAKRRRSDHLALNRSTWMSQSVQVITELVESRIAE  
FRKQGEHISVSVMQSFQVEELDKLSTQAFLDQNLHFAERSTQLLEKQKXEE  
MDNVEIKVVDVTKQVGEESQIAASAEIAGDVLGELITTFHSQHLASYSLGE  
CKSIFGDLLOITSAQKAEARLQELGASQTIYESNATVSARIVLEERQAAEE  
RCNLLLOIGSLINSQAEWMSRLAGKTALIROAVMDSNATFEGSMSYMDGMWAKNEK  
DSRLMEDVVKSRDVLNKKMKDDWATADEHNSIQNTTKSVHAETVRVVDALKDLDTQ

MODLDQFVSRRARSNSVHHEQHVSMENLSTVKSFINIADHGERDRIIDLSEGM  
DNDAKILRESLEFDESRLQALAREDIQSTKREYBPTGDTPEKKRYEYPTDLPT  
AGHETLIAGMDQKPSITTTTSPAKRATMTAVSDIVSPSRSPKVLVSPVSPRNP  
FCTPGPATSRNKSHPMDPSLRVNPVNLGGTSTVFGELSSSTMSMAPSTSVGFHAN  
Query Match 91.6%; Score 17.4; DB 8; Length 60687;  
Best Local Similarity 94.7%; Pred. No. 3.7e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GTCGACGCCGCCGATGACC 19  
|||||  
Db 2647 GTCGACGCCGCCGATGACC 2665  
RESULT 10  
AC068950/c  
LOCUS  
DEFINITION  
Oryza sativa chromosome 10 BAC OSJNBa0041P03 genomic sequence,  
complete sequence.  
AC068950  
AC068950.11 GI:22128701  
HTG.  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Eriharoidae; Oryzaceae; Oryza.  
Buell, C.R., Yuan, Q., Ouyang, S., Liu, J., Gansberger, K., Kim, M.M.,  
Overton II, L.L., Bera, J.J., Tsitrin, T., Krol, M.I., Jarrahi, B.B.,  
Jin, S.S., Koo, H., Zismann, V., Hsiao, J., Blunt, S., Vanaken, S.S.,  
Uttarback, T.T., Feldblum, T.V., Yang, Q.Q., Haas, B.J., Suh, B.B.,  
Peterson, J.J., Quackenbush, J., White, O., Salzberg, S.L. and  
Fraser, C.M.  
Oryza sativa chromosome 10 BAC OSJNBa0041P03 genomic sequence  
Unpublished  
2 (bases 1 to 139614)  
Buell, R.  
Direct Submission  
Submitted (13-MAY-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
Buell, R.  
3 (bases 1 to 139614)  
Direct Submission  
Submitted (31-JUL-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
Buell, R.  
4 (bases 1 to 139614)  
Direct Submission  
Submitted (07-AUG-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org  
On Aug 7, 2002 this sequence version replaced gi:22024565.  
Address all correspondence to:rice@tigr.org  
BAC clone OSJNBa0041P03 is from Oryza sativa chromosome 10  
The orientation of the sequence is from Sp6 to T7 end of the BAC  
clone.  
Genes were identified by a combination of several methods: Gene  
prediction programs including Fgenesh (<http://www.softberry.com/>),  
GENSCAN and GENSCAN+ (Chris Burge,  
<http://CCR-081.msc.edu/GENSCAN.html>), GeneMarkEM (Mark Borodovsky,  
<http://genemark.biology.gatech.edu/GeneMark/>), and GeneSplicer  
(Mihaila Partea and Steven Salzberg, contact [mparta@tigr.org](mailto:mparta@tigr.org)),  
searches of the complete sequence against a peptide database and  
the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>).  
Annotated genes are named to indicate the level of evidence for  
their annotation. Genes with similarity to other proteins are named  
after the database hits. Genes without significant peptide  
similarity but with EST similarity are named as unknown proteins.  
Genes without protein or EST similarity, that are predicted by more  
than two gene prediction programs over most of their length are  
annotated as hypothetical proteins. Genes encoding tRNAs are  
predicted by tRNAscan-SE (Sean Eddy,



<http://genome.wustl.edu/eddy/ERNAscan-SE/>. Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

This BAC overlaps with rice BAC OSJNBa0051D19 (AC023240) and OSJNBa0066108 (AC092489).

There is a small sequencing gap (<500 bp) in this BAC. Closure effort continues.

#### FEATURES

##### Source

1. 139614

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="Genomic DNA"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/chromosome="10"

/map="R2303"

/clone="OSJNBa0041P03"

/note="japonica cultivar-group"

371..8253

/gene="OSJNBa0041P03.1"

/note="similar to second largest subunit of RNA polymerase I GB:AC52850 GI:1621594 (Mus musculus), EST AU184056 from this gene"

join(<371..444,537..654,1422..1506,2098..2190,2722..2834,

3721..3848,3924..4003,4093..4282,4424..4522,4664..4828,

4951..5025,5415..5518,5601..5676,6300..6422,6532..6675,

6989..7201,7373..7606,7810..8253)

/gene="OSJNBa0041P03.1"

join(<371..444,537..654,1422..1506,2098..2190,2722..2834,

3721..3848,3924..4003,4093..4282,4424..4522,4664..4828,

4951..5025,5415..5518,5601..5676,6300..6422,6532..6675,

6989..7201,7373..7606,7810..8253)

/gene="OSJNBa0041P03.1"

/codon\_start=1

/product="putative DNA-directed RNA polymerase I subunit,

5'-partial"

/protein\_id="AAM92814.1"

/db\_xref="GI:22128702"

/translation="VAEAVLKDYFVHLQNNHKNELIIMLKLYAIVDQTSAPDXA

DALQYQVLLPGLITVFLKDLQWLKRLVEEATKNSFDLDFDFLEGV

RGTSVETSGKQVLRIGAEVRFSLKTSAYVGAQSMIKVGVNSQSGDLDP

QRDMTHAERLNFHRYISHERSGSSFAKMTTSVRKLLPSWGLPCVHTPDGE

PCLLVHMTSTCRISSTFNSEKATKDFOKIWSLIARLVGMAQLPRTERTGPPEV

LHLLNGCVISASAKIEEVIPEDLGVYVLSHGAYPGLYLFTWPAFLPAPWE

TRCPDGGGGRNKLFPATHEIHTAILSVVANTPMSDHNQSPNNVQCOMAKQTMG

FCGQALKRTVDKFLHQTPTPIVATYISKYCMDFPSTNAIVAVLYTGYDMED

AMLNKSAVRDMFRGHIFQTECIDLSAKSRDNTVEFFCKNSLSDRTAAISDGLPR

IGENTFPNEQYYSVNNLTGTVPRIKLKSGSEPAIDYVAVNGTDFKDLQKANTLR

VNRPITIGKFSRSGHQGVCSQLWPDIDMPFSANTGRPDLIINPHAFPSRMITAMLL

ESIAAKSLGKGFDTAPPASVYKERSNSIVDELPMLASVGFNYHGTILYSGVFG

TEMKCEITLGPVYVRLSHMYSDFQVTRTCIRIDQITROPIGRGKGGIRFGEMED

ALLAHGASVLLDLHSCSDYHADVCSICSLTATVIKSESOKKAKRMGLIPTVK

FPKPAQACKTSKGMETVAMPYFRYLASELANMKLELRLSNRTEHTTTTSEES"

4861..4900

/rpt\_family="tandem"

6064..6065

/gene="OSJNBa0041P03.1"

/note="about 779 bp of E. coli insertion sequence is1E

sequences deleted"

complement(9020..9102)

/rpt\_family="(CGG)n"

9020..9095

/rpt\_family="tandem"

9046..9102

/rpt\_family="tandem"

9109..13951

/gene="OSJNBa0041P03.9"

/note="predicted by Genemark-ES"

join(<9109..9130,9219..9343,10810..10881,10978..11487,

11595..11909,12003..12059,12235..12561,12641..12847,

12925..13011,13097..13191,13410..13544,13630..13951)

/gene="OSJNBa0041P03.9"

#### CDS

join(9109..9130,9219..9343,10810..10881,10978..11487,  
11595..11909,12003..12059,12235..12561,12641..12847,  
12925..13011,13097..13191,13410..13544,13630..13951)  
/gene="OSJNBa0041P03.9"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAM92816.1"

/db\_xref="GI:22128704"

/translation="MSDSDEYVDLPVSDDEEEDSEDEEKVGRKKAKVHAKOLK

RLQEKDPFYKLEECDEKELFFDDDDPDNDGSAEKHSVPKEPKVITQWQV

DSWCOGAEDKIGSIRSILEARAKACHYGESENNAPKFSVWSSGLDKVHFLYKN

MRIILRELLDAPSGFGKKTYSSELMITQWKRHGRMLRLYLNALHMITELTDEQVA

FIVHRVASDAFAPFALLRIYKALLHNSRGRLAMPVSLFELRDLICIQGSECL

DTSLGIYKAYLVNCKLSKSGSLQHIQIQLGNCVRLYNVDPOSAOHAFVFIQL

AVILRGALTEPGKTSKDKQKESIKPTKKRMEKSYQKYVDWQYIFCLELMTSVVCGC

SSEEDLRPLAYPLTQIIHGVACLVPSARYFFVRLRCVMLNRIAEATGTFIPVSSLL

DLMELKELCGKDPDAVGNALFVSKQVDKTKVTRAFQEBACIFSAVDELAKHLAOWSY

STAFEMSPFLTVRLQNFCKTVKADRFERREIKDLIHOIKASAEFVSSRAGIGRSPND

PAVDSFLOVEKAKSPSLKYVATLHQRSQDMSLDDTSIVTGAESSTFSERLSEAQ

KQDEDDQGEDTIAFSKNLLTEKKTKTPKSKRAHNDHDAVEEDIVEDLLISSD

EDEDDEKMESEDDGSMVEDESDDDFVDPDSQWKKQKKEKSKRNRPQPSKKGSS

TTKRKEKIPHPKKAKH"

9235..9284

/rpt\_family="(GGA)n"

9522..9579

/rpt\_family="tandem"

9539..9570

/rpt\_family="tandem"

14553..14605

/rpt\_family="tandem"

14628..14666

/rpt\_family="tandem"

complement(14849..14883)

/rpt\_family="AT-rich"

15203..16355

/gene="OSJNBa0041P03.11"

/note="predicted by genscan"

join(<15203..15446,15511..16355)

/gene="OSJNBa0041P03.11"

join(15203..15446,15511..16355)

/gene="OSJNBa0041P03.11"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="AAM92818.1"

/db\_xref="GI:22128706"

/translation="MKSPPRAASAPAPAPMPMLSPAGKEPKHFTLQDSSAERDA

DLEAFWRARLVVSSSERVAGDLAYLRRSRSPARGRGGGGGGGGQVQYVCGRGG

VPAVHAGRVAGGVDVARGVGGGGGQACRGDPAGVRRALRUCRRAEGGGGRS

RWRRDGHVARRAGGAGPDAGVGAGVLLRRARRRREAAAGVLDGGEAGEPLRRAR

RAGGPRAAARFPFVAGGAGEGGLLRKRGVPEFGGAAACRGRRRGDAAGAGG

DLQGTIASGCGAPVHAALKRVEVLAELAGAEHDNGMIYMAVPANGELAELEPLQ

ORLVQTPLENTLQGTGHDGGAAGG"

complement(15221..15256)

/rpt\_family="GC-rich"

15517..15661

/rpt\_family="(CGG)n"

complement(15627..15793)

/rpt\_family="(CCGCG)n"

16340..16374

/note="Region with a single high quality sequence

confirmed with a second sequence."

17268..19522

/gene="OSJNBa0041P03.10"

/note="predicted by Jgenesh"

join(<17268..17292,17417..17767,17825..18969,

19490..19522)

/gene="OSJNBa0041P03.10"

join(17268..17292,17417..17767,17825..18969,19490..19522)

/gene="OSJNBa0041P03.10"

/codon\_start=1

Query Match 91.6%; Score 17.4; DB 8; Length 139614;  
Best Local Similarity 94.7%; Pred. No. 3.5e+03;



Qy	Matches	18; Conservative	0; Mismatches	1; Indels	0; Gaps	0;
	1	GTGACGCGCCGATGACC	19			
Db	76086	GTGACGCGCCGATGACC	76068			
RESULT 11						
LOCUS	AP005494	154288 bp	DNA	linear	HTG 05-JUL-2002	
DEFINITION	Oryza sativa (japonica cultivar-group) chromosome 8 clone OSNBA0086M15, *** SEQUENCING IN PROGRESS ***.					
ACCESSION	AP005494					
VERSION	AP005494.1	GI:21699014				
KEYWORDS	HTG; HTGS PHASE2.					
SOURCE	Oryza sativa (japonica cultivar-group)					
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.					
REFERENCE	1	Sasaki, T., Matsumoto, T. and Katayose, Y.				
	Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC clone OSNBA0086M15					
JOURNAL	Published Only in Database (2002)					
REFERENCE	2	(bases 1 to 154288)				
	Sasaki, T., Matsumoto, T. and Katayose, Y.					
TITLE	Direct Submission					
JOURNAL	Submitted (04-JUL-2002) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki, 305-8602, Japan					
COMMENT	(E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)					
	NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.					
FEATURES	* NOTE: This is a 'working draft' sequence.					
	* This sequence will be replaced					
source	* by the finished sequence as soon as it is available and					
	* the accession number will be preserved.					
Location/Qualifiers	1. .154288					
	/organism="Oryza sativa (japonica cultivar-group)"					
ORIGIN	/mol_type="genomic DNA"					
	/cultivar="Nipponbare"					
Query Match	Best Local Similarity	91.6%; Score 17.4; DB 2; Length 154288;				
	Matches	18; Conservative	0; Mismatches	1; Indels	0; Gaps	0;
Qy	1	GTGACGCGCCGATGACC	19			
Db	52622	GTGACGCGCCGATGACC	52640			
RESULT 12						
LOCUS	AF232689	230138 bp	DNA	linear	VRL 18-APR-2002	
DEFINITION	Rat cytomegalovirus Maastricht, complete genome.					
ACCESSION	AF232689	AF046125	U50550	AF077758	U91788	AF133339
VERSION	U87867	AF136519				
KEYWORDS	AF232689.2	GI:20178619				
SOURCE	Rat cytomegalovirus Maastricht					
ORGANISM	Rat cytomegalovirus Maastricht					
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;						
REFERENCE	1	(bases 1 to 230137)				
	Vink, C., Beuken, E. and Bruggeman, C.A.					
TITLE	Cloning and functional characterization of the origin of lytic-phase DNA replication of rat cytomegalovirus					
JOURNAL	J. Gen. Virol. 78 (Pt 11), 2963-2973 (1997)					
REFERENCE	2	(bases 1 to 230138)				
	Vink, C., Beuken, E. and Bruggeman, C.A.					
TITLE	Complete DNA sequence of the rat cytomegalovirus genome					
JOURNAL	J. Virol. 74 (16), 7656-7665 (2000)					
REFERENCE	3	(bases 1 to 230138)				
	Grujthuis, Y.K., Beuken, E., Bruggeman, C.A. and Vink, C.					
TITLE	Rat cytomegalovirus R89 is a highly conserved gene which expresses a spliced transcript					
JOURNAL	Virus Res. 69 (2), 119-130 (2000)					

Betaherpesvirinae; Muromegalovirus.  
1 (bases 61631 to 74978)  
Beuken, E., Slobbe, R., Bruggeman, C.A. and Vink, C.  
Cloning and sequence analysis of the genes encoding DNA polymerase, glycoprotein B, gp18.5 and major DNA-binding protein of rat cytomegalovirus  
J. Gen. Virol. 77 (Pt 7), 1559-1562 (1996)  
96335691  
8757999  
2 (bases 1 to 1511; 229050 to 230137)  
Vink, C., Beuken, E. and Bruggeman, C.A.  
Structure of the rat cytomegalovirus genome termini  
J. Virol. 70 (8), 5221-5229 (1996)  
96357047  
8764031  
3 (bases 74973 to 78970)  
Vink, C., Beuken, E. and Bruggeman, C.A.  
Cloning and functional characterization of the origin of lytic-phase DNA replication of rat cytomegalovirus  
J. Gen. Virol. 78 (Pt 11), 2963-2973 (1997)  
98033207  
9367384  
4 (bases 25540 to 27709)  
Beisser, P.S., Vink, C., Van Dam, J.G., Grauls, G., Vanherle, S.J. and Bruggeman, C.A.  
The R33 G protein-coupled receptor gene of rat cytomegalovirus plays an essential role in the pathogenesis of viral infection  
J. Virol. 72 (3), 2352-2363 (1998)  
98139136  
9499096  
5 (bases 168213 to 176656)  
Beisser, P.S., Kaptein, S.J., Beuken, E., Bruggeman, C.A. and Vink, C.  
The Maastricht strain and England strain of rat cytomegalovirus represent different betaherpesvirus species rather than strains  
Virology 246 (2), 341-351 (1998)  
98327782  
9657952  
6 (bases 98664 to 102381)  
Beisser, P.S., Grauls, G., Bruggeman, C.A. and Vink, C.  
Deletion of the R78 G protein-coupled receptor gene from rat cytomegalovirus results in an attenuated, syncytium-inducing mutant strain  
J. Virol. 73 (9), 7218-7230 (1999)  
99370163  
10438809  
7 (bases 23514 to 27145)  
Beuken, E., Grauls, G., Bruggeman, C.A. and Vink, C.  
The rat cytomegalovirus R32 gene encodes a virion-associated protein that elicits a strong humoral immune response in infected rats  
J. Gen. Virol. 80 (Pt 10), 2719-2728 (1999)  
20037837  
10573166  
8 (bases 194997 to 204137)  
Beisser, P.S., Kloover, J.S., Grauls, G.E., Blok, M.J., Bruggeman, C.A. and Vink, C.  
The r144 major histocompatibility complex class I-like gene of rat cytomegalovirus is dispensable for both acute and long-term infection in the immunocompromised host  
J. Virol. 74 (2), 1045-1050 (2000)  
20091365  
10623772  
9 (bases 1 to 230138)  
Vink, C., Beuken, E. and Bruggeman, C.A.  
Complete DNA sequence of the rat cytomegalovirus genome  
J. Virol. 74 (16), 7656-7665 (2000)  
20366325  
10906222  
10 (bases 1 to 230138)  
Grujthuis, Y.K., Beuken, E., Bruggeman, C.A. and Vink, C.  
Rat cytomegalovirus R89 is a highly conserved gene which expresses a spliced transcript  
Virus Res. 69 (2), 119-130 (2000)



JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	metabolites Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001) 21477403 11572948 2
TITLE	Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	Nat. Biotechnol. 21 (5), 526-531 (2003) 22608306 12692562 3 (bases 1 to 299175)
TITLE	Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kushida,N., Director-General of Biotechnology Center, Shiba.T., Sakaki,Y. and Hattori,M. Direct Submission
JOURNAL	Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0065, Japan (E-mail:biocente.go.jp, URL: <a href="http://www.bio.nite.go.jp/">http://www.bio.nite.go.jp/</a> , Tel:81-3-3481-1933, Fax:81-3-3481-8424)
COMMENT	This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3). Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa. *1 Kitasato Institute for Life Sciences, Kitasato University *2 National Institute of Infectious Diseases *3 The Kitasato Institute *4 National Institute of Technology and Evaluation *5 School of Science, Kitasato University *6 Institute of Medical Science, University of Tokyo *7 RIKEN, Genomic Sciences Center Following url is also available. <a href="http://avermitilis.ls.kitasato-u.ac.jp">http://avermitilis.ls.kitasato-u.ac.jp</a> .
FEATURES source	Location/Qualifiers 1..299175 /organism="Streptomyces avermitilis MA-4680" /mol_type="Genomic DNA" /strain="MA-4680" /db_xref="taxon:227882" /note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermitinus" complement(155..550) /note="SAV445" Tn4561, truncated /codon_start=1 /transl_table=1 /product="putative transposase" /protein_id="BAC68155.1" /db_xref="GI:29604084" /translation="MAGPPRSATRSRTADRESAHPADPEFGYRRQIKVQANLQ EGHLLARRAGQLYQYQDMEDQIGALGLVNLALVFNTRYMDALTQLRADGFEVR DEDVARLSPFWRHHINVIGRYSFQLPDL" complement(474..740) /note="IS630 family, truncated SAV446" /codon_start=1 /transl_table=1 /product="putative transposase" /protein_id="BAC68156.1" /db_xref="GI:29604085" /translation="MPTQNAQPSSTVTAARAVSTAGSVYNNAPTSPVMAPPSPGRRR EPGSLGAVRAYDIVRMILSRDGRPTPLGDALITYGGRRLCTSCA" 766..987 /note="SAV447" /codon_start=1
CDS	1..299175 /product="putative transposase" /protein_id="BAC68157.1" /db_xref="GI:29604086" /translation="MUVVAVGCALTSVPTAGGVVAVPCVRRRLTDOEGKLOQIVRRG SASSVPRPRAMLLASAGNRVPVIAQLV" complement(1954..2571) /note="SAV448" /codon_start=1 /transl_table=1 /product="putative TecR-family transcriptional regulator" /protein_id="BAC68158.1" /db_xref="GI:29604087" /translation="MFTSRRARERANTRERIEAALHVLTEGIALIRIRIATDVE YSAPVYQHFANKDALVILVAHGHRLMLSEFSQAQEPDTRRMTRIASQVYRFAGE HPLFQVMDNPMVDADERRVAEPVIDVLKELITSWAADHVVLADFQACEILWGTL YGIASLYSGVGNDRARYLAEQALHTILLGWRTETPANGRPASS" 2694..3098 /note="SAV449" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="BAC68159.1" /db_xref="GI:29604088" /translation="MIYHNRVTMKATPEQIEGVLESWRNQGRSNPAIKSFVQGRN HGSDYTYGAVFVVEHLDGLFAVLTHTPTTCTDHLGLHLVERLEIFDVSDDDPDLANK IQELHRRNELNPQIAGMLADVPTTYSGVDD" 3127..4029 /note="SAV450" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="BAC68160.1" /db_xref="GI:29604089" /translation="MAKHVVIIGAGSIGTNNVARLLAEGGESVRIIVTSGSGPEHPLVDR VADADSPRLTLELSRGAGVIYHCANPPSYTWERLLPPLQTAATPAKANDAVLALT GSLYAYGPPGGRMNEHTPMATGHKRLRRMWEQALAGGIRITVVRGSDYIGKDN GIYLLIKSALQKAAWITGHLDMPHTTFNGDMAQALITLASEERAKGRPWHPVPSA PAVITRELRARYAVAAGRPPIKLIQIPRSVTRTAGLVFIAREMAEMDYQWVAPPMD ATETTFGLTALDLDVAVRNVG" complement(4130..4498) /gene="bidg" complement(4130..4498) /gene="bidg" /note="SAV451" /codon_start=1 /transl_table=1 /product="putative anti-anti-sigma factor" /protein_id="BAC68161.1" /db_xref="GI:29604090" /translation="MTDAFHIDVPHPTDAALAVIALSGEPDIAAPAVRARALELIANG HPDLVADLSGVTFCDSGLGALVIGWCADKADGSLTAAIPDRLSRLLSVTGMDAFL PAYSSADAALAAARQGNRTTA" 4635..5156 /note="SAV452" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="BAC68162.1" /db_xref="GI:29604091" /translation="MFAVXESRVRVGEIAATPTGEALSLVHLGSLGVRVESDDA TARPDGLDFAELDPARGEWHTLALVRPVEGEEAGLFPFGAPGSGPARLCWRE SAPRITDIEDPEVHRTLRSLEDLGAHLRPAAPRTHGGRRCRCLPHGVFPRLLAAV LVHGTAAAGCTPR" 5183..5665 /note="SAV453" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="BAC68163.1" /db_xref="GI:29604092" /translation="MVTAVRARAGSAWQACRGARRSCHSHSEGVRMDEKEFLRTV AERTGLSRQEAADLTWATLETTLARLSPGEARDLTLPGLAEAVRRTCTDRIERFG

HHDSVRRVAERTMLKEEADRGVTRVLAVALREAISEKEFNDLMSQLGTDPSQALASAG  
 " 5810. .6964  
 /note="SAV454"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC68164.1"  
 /db\_xref="GI:29604093"  
 /translation="MIMTADVPETSSGSRVSRVALVADPGLPSEIARALAPRLPDHLRR  
 AVGARQVQDMTAPLSVSCADVSDIABAVRSHLDGAWDVAIFLTPDPRRLVLP  
 TSVEDVTALRALISLALGVRLRRVQAVDVVRELVAHEETPRHPVEIGRPPE  
 PVDTRPFEQQRVWFGLGHTFLVAGMYANRPHRLFGSLRSLAGVAPATVFLF  
 INSATNLATAGLVCHAVITAGSAVALTLLIIVDQWLGKRGTHLPHRHFLYFYNLV  
 TLVTIALAVVLAALFATLVAVSFLLDASVLRKFGKRGVGVGDYFLFLAWFITAIAM  
 IGGAFGTSLGEGKVSADAYGRQRQRMLRQMPRPERGERGPDPTDG"  
 Complement (6986. .7498)  
 /note="SAV455"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC68165.1"  
 /db\_xref="GI:29604094"  
 /translation="WNPVAPASGIBQRPQWMLDYLWCSDIIDLAEGLEQRPLV  
 ATVRLAEGERVLAVGPAQRTVTRALGDRYMHQNAVAFGSPAFVLGSMASMATSLG  
 NSARRAANDAPRCGARTGLHLVDLVDHGTMLRWSMELGTSSSTRPSSALRALI  
 ELVDLHVIAL"  
 7548. .8189  
 /note="SAV456"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="BAC68166.1"  
 /db\_xref="GI:29604095"  
 /translation="WATLLRRVCPHEGVRIAAVAGFVLSRRPNTASVYRCVAVGPOS  
 EDREVEGSRTPVASRLRPLRRRLSVRCPRAVPSQVARSFRFGTRFLGDDVL  
 VTQRDHALHARVRIIDVHPGLGRGAARFVRIATHSLGGVIAVDWATANVPLWTS  
 LVTFGSQATFHHPQVPGDLLCOMIAFRMRKIQEABTGALPPTGFIHAGQGW"  
 Query Match 91.6%; Score 17.4; DB 1; Length 299175;  
 Best Local Similarity 94.7%; Score No. 3.3e-03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCGAGCGCCGCGATGACC 19  
 |||||  
 Db 187408 GTCGAGCGCCGCGATCACC 187390  
 RESULT 14  
 AP005949 299850 bp DNA linear BCT 12-SEP-2003  
 LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section  
 DEFINITION 15/31  
 ACCESSION AP005949 BA000040  
 VERSION AP005949.1 GI:27352050  
 KEYWORDS  
 SOURCE Bradyrhizobium japonicum USDA 110  
 ORGANISM Bradyrhizobium japonicum USDA 110  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Bradyrhizobiaceae; Bradyrhizobium.  
 REFERENCE 1  
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,  
 Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K.,  
 Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.  
 and Tabata, S.  
 Complete genome sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110  
 DNA Res. 9 (6), 189-197 (2002)  
 MEDLINE 22484998  
 PUBMED 12597275  
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,  
 Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K.,  
 Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.  
 and Tabata, S.  
 Complete genome sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110  
 DNA Res. 9 (6), 189-197 (2002)  
 MEDLINE 22484998  
 PUBMED 12597275  
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,  
 Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K.,  
 Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.  
 and Tabata, S.  
 Complete genome sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110 (supplement)  
 DNA Res. 9 (6), 225-256 (2002)  
 MEDLINE 22485002  
 PUBMED 12597279  
 REFERENCE 3 (bases 1 to 299850)  
 AUTHORS Kaneko, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research; 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: kaneko@kazusa.or.jp)  
 URL: http://www.kazusa.or.jp/rhizobase/  
 Tel: 81-438-52-3935 (ex. 2398), Fax: 81-438-52-3934)  
 FEATURES  
 Location/Qualifiers  
 1. 299850  
 /organism="Bradyrhizobium japonicum USDA 110"  
 /mol\_type="Genomic DNA"  
 /strain="USDA110"  
 /db\_xref="taxon:224911"  
 18. 11400  
 /gene="cysS"  
 18. 11400  
 /gene="cysS"  
 /note="ORF ID:blr3790"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cysteineyl-tRNA synthetase"  
 /protein\_id="BAC49055.1"  
 /db\_xref="GI:27352051"  
 /translation="MELRLYDLTKRFTVFLPDANNVYVCGTVYDFAHGNARP  
 VIFVDVPLRLHLYGEAHVKNRITVDKINDRAADFPLGLPNEAIREVTEQTG  
 KQHPADVADGALRPSVEPRATEHIGEMREIIRLVAGGFAYVAGDHVLFSPQANAA  
 NSTLPRYALSNSRLEDMVAGARVAVPYKGSITDFVLWPKSPGSPWSPSPAGIAAQ  
 GFPWIECSANAKHGLBTFDIHGGIDILVPHHENEVAQCCAFHERRMANYMEN  
 GFLVSEKSKSLGNFITHELLADWPGEVLRKLNKTHYRSPIDMTWKSLESAKT  
 LDYRYAADVDFGPAAVVEFLDDNLNTSLAALHGLRNSDVSLAGSLRLGLFL  
 SESAAQWEDRKQKASGVDAAGEVERLISERTAAAGRKDFKESDIRDLAAMGVAIKDS  
 KEGITWERSK"  
 1504. 2007  
 /gene="blr3791"  
 1504. 2007  
 /gene="blr3791"  
 /note="ORF ID:blr3791"  
 hypothetical protein  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC49056.1"  
 /db\_xref="GI:27352052"  
 /translation="MVKTWQTLPRGLRPFLDDVPVLAALFVASIELTGTEDYEA  
 QQQAWYEAASEEFGRKLAADLTLTATLEGPVGFASLQGDYIRKLYVHPVAVGGGI  
 ATMLDALBKLACGGKAKSLSDASDNAQGFPAKRGYVAMQNSITVDNEWLANMTK  
 KTLGAPO"  
 2186. 3799  
 /gene="blr3792"  
 2186. 3799  
 /gene="blr3792"  
 /note="ORF ID:blr3792"  
 /codon\_start=1  
 /transl\_table=11  
 /product="2-isopropylmalate synthase"  
 /protein\_id="BAC49057.1"  
 /db\_xref="GI:27352053"  
 /translation="MSSAQAQRERLYVFTTLRDGQNGVDTLTKQIIIAAMLDL  
 GIDYGVGGPGANPTDTEFGSKPKLNHARFTAFGMRAGRSVNSNDCEPVGALLAKA  
 DAICFVAKSGSAQVRVALETKEENLASIRDSVAAKAGREVMLDCEHFFDGYKEDS  
 SPALACAKAAYEAGARWVLCDTNGTMEVEAIVTEVTRI PGHVGFIHAINDTPEQ  
 AVANLAAVRAGARQIQGTNLGLGRCGNALCSLPTLKLKQEFADAFEI GVTPEKL  
 ATLVKYSRTLDDMLNVRNHAAYGESAFVTKTGHASAVLKDPQTYBHVLPGLVGN  
 HRKVLSQDQGRSNVIAELDRAGIAYEKSDPKLTRIVELKEREAGAYVESANSGD  
 LLARRTLGKVPHYFEVEQFDVNVQEQINALGERVTVALAVKVDVAGEHLISNAESNG

Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M.  
 and Tabata, S.  
 Complete genome sequence of nitrogen-fixing symbiotic bacterium  
 Bradyrhizobium japonicum USDA110 (supplement)  
 DNA Res. 9 (6), 225-256 (2002)  
 MEDLINE 22485002  
 PUBMED 12597279  
 REFERENCE 3 (bases 1 to 299850)  
 AUTHORS Kaneko, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research; 2-6-7  
 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: kaneko@kazusa.or.jp)  
 URL: http://www.kazusa.or.jp/rhizobase/  
 Tel: 81-438-52-3935 (ex. 2398), Fax: 81-438-52-3934)  
 FEATURES  
 Location/Qualifiers  
 1. 299850  
 /organism="Bradyrhizobium japonicum USDA 110"  
 /mol\_type="Genomic DNA"  
 /strain="USDA110"  
 /db\_xref="taxon:224911"  
 18. 11400  
 /gene="cysS"  
 18. 11400  
 /gene="cysS"  
 /note="ORF ID:blr3790"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cysteineyl-tRNA synthetase"  
 /protein\_id="BAC49055.1"  
 /db\_xref="GI:27352051"  
 /translation="MELRLYDLTKRFTVFLPDANNVYVCGTVYDFAHGNARP  
 VIFVDVPLRLHLYGEAHVKNRITVDKINDRAADFPLGLPNEAIREVTEQTG  
 KQHPADVADGALRPSVEPRATEHIGEMREIIRLVAGGFAYVAGDHVLFSPQANAA  
 NSTLPRYALSNSRLEDMVAGARVAVPYKGSITDFVLWPKSPGSPWSPSPAGIAAQ  
 GFPWIECSANAKHGLBTFDIHGGIDILVPHHENEVAQCCAFHERRMANYMEN  
 GFLVSEKSKSLGNFITHELLADWPGEVLRKLNKTHYRSPIDMTWKSLESAKT  
 LDYRYAADVDFGPAAVVEFLDDNLNTSLAALHGLRNSDVSLAGSLRLGLFL  
 SESAAQWEDRKQKASGVDAAGEVERLISERTAAAGRKDFKESDIRDLAAMGVAIKDS  
 KEGITWERSK"  
 1504. 2007  
 /gene="blr3791"  
 1504. 2007  
 /gene="blr3791"  
 /note="ORF ID:blr3791"  
 hypothetical protein  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAC49056.1"  
 /db\_xref="GI:27352052"  
 /translation="MVKTWQTLPRGLRPFLDDVPVLAALFVASIELTGTEDYEA  
 QQQAWYEAASEEFGRKLAADLTLTATLEGPVGFASLQGDYIRKLYVHPVAVGGGI  
 ATMLDALBKLACGGKAKSLSDASDNAQGFPAKRGYVAMQNSITVDNEWLANMTK  
 KTLGAPO"  
 2186. 3799  
 /gene="blr3792"  
 2186. 3799  
 /gene="blr3792"  
 /note="ORF ID:blr3792"  
 /codon\_start=1  
 /transl\_table=11  
 /product="2-isopropylmalate synthase"  
 /protein\_id="BAC49057.1"  
 /db\_xref="GI:27352053"  
 /translation="MSSAQAQRERLYVFTTLRDGQNGVDTLTKQIIIAAMLDL  
 GIDYGVGGPGANPTDTEFGSKPKLNHARFTAFGMRAGRSVNSNDCEPVGALLAKA  
 DAICFVAKSGSAQVRVALETKEENLASIRDSVAAKAGREVMLDCEHFFDGYKEDS  
 SPALACAKAAYEAGARWVLCDTNGTMEVEAIVTEVTRI PGHVGFIHAINDTPEQ  
 AVANLAAVRAGARQIQGTNLGLGRCGNALCSLPTLKLKQEFADAFEI GVTPEKL  
 ATLVKYSRTLDDMLNVRNHAAYGESAFVTKTGHASAVLKDPQTYBHVLPGLVGN  
 HRKVLSQDQGRSNVIAELDRAGIAYEKSDPKLTRIVELKEREAGAYVESANSGD  
 LLARRTLGKVPHYFEVEQFDVNVQEQINALGERVTVALAVKVDVAGEHLISNAESNG

```

PVALDVALRKDLGKYOKYIEGLKLDYRVRILNGGTGAVTRVILSEEDENGDSWTTV
GVSNIIDASFOALMDSVIYKLVKSGAPA"
3930. 4307
/gene="b1r3793"
3930. 4307
/gene="b1r3793"
/note="ORF ID: b1r3793
putative lyase"
/codon_start=1
/transl_table=11
/protein_id="BAC49058.1"
/db_xref="GI:27352054"
/translation="MIDHISVGVSDLDLRSATFVEATLAALGLRLVWRPRTVFGKAY
PEWNLREGMPVPVPSGVHICLRKATTAEDVAFHAAALSAGASDGPAGIRPHDRV
RYTAAFPVDPGNRIEAVTFPAE"
complement(4489. .5103)
/gene="b1r3794"
complement(4489. .5103)
/gene="b1r3794"
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49059.1"
/db_xref="GI:27352055"
/translation="MAPMSTIKTVVCVCGSGPGTNPRFTGAKAFKALAEENIRLVY
GGSLGMSVATSLVDNGGTGIIPEFLRENAITRVOEMIVTPDMHERKLMFE
RSDAFVALPGVGLLELVEQLTKQLGRHAKPVLANIINFWEPLFSLSHMRQTEF
IRAGLSVILKADRVBEILPKLKAQAVALAEKELAPEVARKL"
5295. 7250
/gene="b1r3795"
5295. 7250
/gene="b1r3795"
/note="ORF ID: b1r3795"
/codon_start=1
/transl_table=11
/product="ABC transporter HlyB/MedA family"
/protein_id="BAC49060.1"
/db_xref="GI:27352056"
/translation="MNQPHSPGADVPDPAGGLPERATLMTGLAHLWPIYWPQGRFDL
KYRVWMLVLLAAKLITLVPPFKWATDALGTANTAPVLPDNLHLMVIAPELLTA
SYGATRIALMTQRLGFLARVAMHAKLATITFTHMELSLRFLERKTGLTRV
LERGREGIVRVILQIPTAIVESLLVLLWQPDWRYVVAITLVAVVWYTVI
ATEWREIGIRKMDSDTEANTKAIDSLINVEYKYFGAAREAOYDRSVARVEESSV
QAYSLAVNTGQAVITLGLTATMLMCAIGVNGTNGDDEVUNAMQIQPLNF
MGWYREIKQAIDIERKMGVIGREAEIKQDPQPLVISAGIVREDVFAPEPAP
ILKGISFEVPKQVAIVGSPGAGKSTISRLFLYDISGRILIDQDIFREYQASL
RASICMPQDTLVFNDRINRYRWDAEDAEVEEARLAQIDNFIRMADPKGYEQV
GERGLSGGKQVAVARTVLKAPPIVLDEATSDLTHTHEIIOGALDRVAKNRTS
LVIAHRLSTIVADEIIVLDQRIAERGTHAKLLAQLAAGLVASMMWRQEAAREKLA
KWADTSAPNRQPPVEDATTPAAE"
7334. 8032
/gene="b1r3796"
7334. 8032
/gene="b1r3796"
/note="ORF ID: b1r3796
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49061.1"
/db_xref="GI:27352057"
/translation="MSIIDSITORQIPPIHKEGYPIGGFALASILFLWSPILGWIGT
ILTVKCALFFDPVVRVTPVREGLVSPADGVSMTVAPLPAELGLGDRPLPRISVFM
SVFNCHVNRSPDIAGVRDRIAYRPGLFINAEGLKASDNEKNSLVITTPRIGVQVIA
GLIAKRVCFVREGQACAGERFGLIRFGSLDYLPVGTIKALVSEGTALAGETILA
DIAGDPSRAYRAN"
8120. 9001
/gene="psaa"
8120. 9001
/gene="psaa"
/note="ORF ID: b1r3797"
/codon_start=1
/transl_table=11

```

```

/product="phosphatidylserine synthase"
/protein_id="BAC49062.1"
/db_xref="GI:27352058"
/translation="MTPVDIKDDPVRRRRFRPIVRMLVBNVITLALCAGLTIRLS
TEGRMSLAYAIVFAAALDGDIGRIARNIKQSKGKAGLDSLADGVAGPLMYE
WQLHELGNAGWIAVNFALISGRLRARNATNDENKEPAPANFTGVPAPAGATVIL
LPIYVAFDLGRAPAAVTAATLLIAFLMWSRKLPVSGTKRMVRPPELVLPFAFVAVI
VFATLAIYPMWHSIGTVLYLLALPLGYSYRQDQARDAATAATPAGBISPPSAPTLL
ANLSEPPHDDDRPGRHLH"
9175. 9942
/gene="b1r3798"
9175. 9942
/gene="b1r3798"
/note="ORF ID: b1r3798
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49063.1"
/db_xref="GI:27352059"
/translation="MPAASRQAANLGENAVTDTATGPLPASVLEALGRYDPTTCNA
MEIVAPERRLIGYTTKQLVCPFPDPPIPVYGVARTVAIRSVLKSSLPABEQKRRIEY
EYVGTGHPRISVIOIDIDPDVYGAFWGVQSNVHKALGCLGVTIDGSIIDPQWAP
GFQALAGSIGPSHAWHAEISFGGEVVRVAGMTVKSDLLIHADQHGAIIVPLDVAALKE
AAELCGRRETPILEIARGSPDFSLKLEKALKRSABIH"
10314. 11171
/gene="b1r3799"
10314. 11171
/gene="b1r3799"
/note="ORF ID: b1r3799
probable oxidoreductase"
/codon_start=1

```

Query Match 91.6%; Score 17.4; DB 1; Length 299850;  
Best Local Similarity 94.7%; Pred. No. 3.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGACC 19  
|||||  
Db 250812 GTCGACGCCGCCGATGACC 250830  
|||||

RESULT 15  
AE017108  
LOCUS  
DEFINITION  
Oryza sativa (japonica cultivar-group) chromosome 10, section 62 of  
77 of the complete sequence.  
ACCESSION  
AE017108 AE016959  
VERSION  
AE017108.1 GI:31432797  
KEYWORDS  
SOURCE  
ORGANISM  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Erihartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 301666)

REFERENCE  
AUTHORS  
CONSTRM  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

The Rice Chromosome 10 Sequencing Consortium  
In-depth view of structure, activity, and evolution of rice  
chromosome 10  
Science 300, 1566-1569 (2003)  
2 (bases 1 to 301666)  
Buell, C.R., Wang, R.A., McCombie, W.R., Messing, J. and Yuan, Q.  
Direct Submission  
Submitted (05-MAY-2003) The Institute for Genomic Research, 9712  
Medical Center Dr. Rockville, MD 20850, USA  
This is the pseudomolecule for rice chromosome 10, which was  
constructed by resolving discrepancies between overlapping BACs,  
trimming the overlap regions, and linking the unique sequences to  
form a contiguous sequence. Genes in individual BAC clone were  
identified by a combination of several methods: Gene prediction  
programs, searches of the complete sequence against a peptide  
database and EST databases. Genes with similarity to other proteins  
are named after the database hits. Genes without significant

peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).

FEATURES  
 source  
 i. .301666  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="10"

misc\_feature  
 i. .131983  
 /note="Chromosome Sequence Derivation: nucleotide sequence in this region was derived from BAC clone OSUNBa0051D19 (GB:AC023240)."  
 <2255. .>2896  
 /locus\_tag="OSUNBa0051D19.24"  
 /note="predicted by Genemark hmm"  
 <2255. .>2896  
 /locus\_tag="OSUNBa0051D19.24"  
 <2255. .>2896  
 /locus\_tag="OSUNBa0051D19.24"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAP54385.1"  
 /db\_xref="GI:31432798"

gene  
 /translation="VAGAWTMSRRRLSHALLLVAILPNLAALAAVAGGGGGGFFP  
 HSLQSSCPGSAITGSDCNADNGECQDGVQVMTYACPPVAAAGTGAALL  
 FNSADGCGAGSCPTGRFYDDQQLVALSTGFWGDRSCERKDWIRASGGASVTA  
 MVVDCDORGCDSDHNEFPFCRNIVDGSAPVWDALGNLKDGEAQITWSDA"  
 complement(<3441. .>4021)  
 /locus\_tag="OSUNBa0051D19.23"  
 /note="similar to metallothionein GI:3860191 (Silene vulgaris)"

gene  
 complement(join(<3441. .3698,3833. .>4021))  
 /locus\_tag="OSUNBa0051D19.23"  
 complement(join(<3441. .3698,3833. .4021))  
 /locus\_tag="OSUNBa0051D19.23"  
 /codon\_start=1  
 /product="putative metallothionein"  
 /protein\_id="AAP54386.1"  
 /db\_xref="GI:31432799"

gene  
 /translation="MAALEASGHWGFFRVGLRCGEAPDCHRELFWVSGMGEVLA  
 AVDLTVGVEKMAAKFAAGEGEGESPPCTGDDGAEVAGDARDGRRRCRSGKK  
 HGCRTHAVSGVGVGEGAGESLCTWELRERKRSGEGEEKVRPA"  
 complement(<5481. .>7359)  
 /locus\_tag="OSUNBa0051D19.17"  
 /note="similar to Tam3 transposase GI:16063 (Antirrhinum majus)"

gene  
 complement(join(<5481. .6777,6859. .7041,7133. .>7359))  
 /locus\_tag="OSUNBa0051D19.17"  
 complement(join(<5481. .6777,6859. .7041,7133. .7359))  
 /locus\_tag="OSUNBa0051D19.17"  
 /codon\_start=1  
 /product="putative Tam3-transposase"  
 /protein\_id="AAP54387.1"  
 /db\_xref="GI:31432800"

gene  
 /translation="MGPTLQAQSSLRFSNVEGKGVSDVDTYFDPVARRRELAAMIILH  
 EPLGIVDHIGPRRVMAQLPFKAVTNTIRKDIMTYEKKRALAYMAGASRVG  
 ITDLWTNOKRGYMAITAFHIDSDWTLRSI IMRFIVVPAPTAEVICHYEALVE  
 NMLDKISTLIDNTDKVLSLKKIGRKLMLGKALHMECAHILNLIVDGL  
 DVKQIAKRVSAFWATPKRVKEFEIAKVKVKNLGLDCTKRWNTYRMLSL  
 TALCARADRAIRVATFDCAPSEEEWAFASVVDVRLKFNDFINTAFVSGSYNTSMI  
 QLLKICEKEQIRKAWAGCSTIEMSVMTQKFDKWKDIOGPMGLATLDPRFKID  
 YLLGFIETIGSSSECATKVAEVDYILMKYEVEDDEDNTESSAPPLVNSDLSL  
 SITAVTGRPPAIVKVELDRLDYELVNTNFNFKLDMWKVAGTSFTLRKVARD  
 IPALPVSTVASSAFSTGRVLSRSLTFELLEALMCSQDLRNKRYGTYDVLFIH  
 HA"  
 complement(<8562. .>9337)  
 /locus\_tag="OSUNBa0051D19.16"  
 /note="predicted by Genemark hmm"

mRNA  
 CDS  
 gene  
 mRNA  
 CDS  
 gene  
 mRNA  
 CDS  
 gene

complement(join(<8562. .8938,9148. .>9337))  
 /locus\_tag="OSUNBa0051D19.16"  
 complement(join(<8562. .8938,9148. .>9337))  
 /locus\_tag="OSUNBa0051D19.16"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAP54388.1"  
 /db\_xref="GI:31432801"  
 /translation="WAMGRSPVPQPPFPHILPLGDGDGNCVSMGNQWERKYPGSLGA  
 AGRHVPSPSPSVEARYKSVRALRAAPAVASAPAGRHARRVAVRSSAPRPPVP  
 AGRHVPSPSPSVEARYKSVRALRAAPAVASAPAGRHARRVAVRSSAPRPPVP  
 GKKGENDESTRRPAARVRHGFAFTAKL"  
 <11512. .>14348  
 /locus\_tag="OSUNBa0051D19.22"  
 /note="Predicted by Genemark hmm"  
 join(<11512. .11580,11683. .11742,11826. .11894,12814. .13188,  
 13280. .13337,13752. .13847,13959. .14017,14109. .14175,  
 14263. .>14348)  
 /locus\_tag="OSUNBa0051D19.22"  
 join(<11512. .11580,11683. .11742,11826. .11894,12814. .13188,  
 13280. .13337,13752. .13847,13959. .14017,14109. .14175,  
 14263. .>14348)  
 /locus\_tag="OSUNBa0051D19.22"  
 /codon\_start=1  
 /product="hypothetical protein"  
 /protein\_id="AAP54389.1"  
 /db\_xref="GI:31432802"

gene  
 /translation="WAPAAKRNIGKAPIVDDSDSCDGLSRRLDDIIEIFRKEV  
 VDKKEILNRKSCSHAFRYAKTVCDIFGLPSGGLFEGKDFEAGKEYILSMYGLSCLP  
 SVRFQDQIKKEPLNEKVIITSLFIALACLPNSLIPSTKLTITIFQDVNNUNRY  
 DWSKFTYDSNMKMKFVNTSLGGLYMAVYLDNVDFGDNNSNIPHQNVIEE  
 NIENYVTSKADDDIGISAGISAQVHPVNDVLPSPHPTIHNAFHGNEFAFVKSVEL  
 NDSRQDDDLNFVTPQVGNANHSKQSDVLLDGLP"  
 complement(<26683. .>31718)  
 /locus\_tag="OSUNBa0051D19.15"  
 /note="similar to gag-pol polyprotein GI:4234852 (Zea mays)"

gene  
 complement(join(<26683. .30592,30833. .31004,31574. .>31718))  
 /locus\_tag="OSUNBa0051D19.15"  
 complement(join(<26683. .30592,30833. .31004,31574. .>31718))  
 /locus\_tag="OSUNBa0051D19.15"  
 /codon\_start=1  
 /product="putative gag-pol polyprotein"  
 /protein\_id="AAP54390.1"  
 /db\_xref="GI:31432803"

gene  
 /translation="MASAPAAVAMAMIFLLAALSTHLLASSLRPGAVAGGACRPSGYL  
 PGKSGRAIMFLGSAARDYSLFNTYFIPTNNTGDKXETPTVNTNGNTASNSGG  
 PFLGSPVSVGFAALPHADGSKYNTKRWKTRALLMLTAMQCFYVSGRKRSEPLGPEE  
 EAKFEASDCLFRGALISVLADNIVDYHMPSGKMDRALEAKFGVSGSESLYMEQ  
 FDYKXNDLRSVVEQAEHOMLAKELENNCLPDKFVAGGIIAKLPSSWSPATSLX  
 HKRQFSPDLIGSLGVEBEKAKDVRGKVEGSSANVQKKNPHSNKKKAPDV  
 KPAATNFKKSGKAKGDFVCGSGHWKDCPERKORSANMIISEGGTSGYGI  
 LPTLVFHSFPMVDVTGANIHVCADI SLFSSYQVGRGSSLLMGNSLAAGVGTVD  
 LKFTSGKTVQLKNVQHPSPKIKNLVSGSLLCREDFLVPSNKKVYSGTFTVGGYD  
 SGLPRFLSNDKNHNAVHISENDSNVHSLCHVNFQMTLANNLSLTPKFTLV  
 KQSKCHTQVQKQPKPKHKASEARNLAPELHSDLCENMGVLTGKKKLYFTLLDDC  
 TRFCYVLLTKDEALHPFKIYKAVENQLEPKIKELSDRGGVFSNEFAFCFEFG  
 IHEMTPTPSQSGVAKERNLTETEMVNADLTAGLSKEMWGEAVLTACHVNLKIPM  
 KHKEVTPFEWERKLNLSYLTWGLCAKVNYPFAKKRKLGPKTVDCCVFLGTAIHSV  
 YKPLVNSGPDHVTITLESADTFFENFPMKYTPSTSKETVMPHEFAPISHND  
 QPEENPEINDVTDRKSRQVAKSGDDYIVLVDDTPTTIEEYSPADYWEKA  
 VQSEMDISNGTWEVPERPGKPGKPFVKKLPDGTIEKYKALVAKGYQKE  
 GDFDTYSYVARLTIRVLLAAASHGLVHQMVKTAFLNGELEEBEILYMDQPGYV  
 LBGQEGYCKLLSLYGLKQAPKQHEKPDFTLTSGAFVNAEDCVVYVGGSEVI  
 LCIVDDIIFIGLSNVIEEKDYLSKSFEMKDLGEADVILNKLQREGGCTILVQS  
 HYFVKSLSFGSDCKPAPTPYPSVLLRNRRIRADQRYSQIIGSLMILASATRPD  
 ISFVAKSRFVSNPDGDRALERMVLYLQKTSYGYHYTGYPKVLEGSYNSWD  
 ADEIKATGYFTLGGAVSWKSCQTLITRSTMEALTALDTAVEAWRELLMDL  
 PVYKRPVAILANCONQVIVIKVNSKONKMSRHIKRLKSVRKQKNSGVIALDYVQ  
 TARNLADQFTKGLPCNVIDASREMLIPT"  
 complement(<23257. .>35641)  
 /locus\_tag="OSUNBa0051D19.14"  
 /note="predicted by Genemark hmm"

```
mRNA      complement(join(<32357..32362,32539..32932,33018..33121,
35012..>35641))
/locus_tag="OSUNBa0051D19.14"
CDS        complement(join(32357..32362,32539..32932,33018..33121,
35012..35641))
/locus_tag="OSUNBa0051D19.14"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAP54391.1"
/db_xref="GI:31432804"
/translation="MAANVKVVLALLQLMSLHVVHGGDNGVSAVATGKHPKPK
QTGGGGGGGCGHCHGKAGKCNRAHSGDCVTRRYPQRCPPVPPVSAAPTAT
LQNSFARGDGGGSCDGRFHPDPTAMVALSSGMLRLDCAKRCNMRIRVAAGNCRS
ALHVVDCDSVNGDCAHNFPEPCPNVDVDSVAVKALDGEVGERFASLLSTS
PTLVGSDAASCHASGVYVHGNETFTAILTLKVPDHGEDDGGVPCSDMRFRN
TELVAUSSWLRGGRRCHRRIRFVAVAGRSGSSVAVRVDDCDSDVNGCRRED
GFAPPCRNNAVGSPPVWEKLGNASVEFEVWNSCLKNAKEL"
gene       complement(<39112..>39813)
/locus_tag="OSUNBa0051D19.13"
/notes="similar to secretory protein GI:5669008 (Triticum
aestivum)"
mRNA      complement(<39112..>39813)
CDS        complement(<39112..>39813)
Query Match      91.6%; Score 17.4; DB 8; Length 301666;
Best Local Similarity 94.7%; Pred. No. 3.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTCAGCGCGCGGATGACC 19
|||||
Db 193940 GTCAGCGCGCGGATGACC 193958

RESULT 16
LOCUS      BX640451 343473 bp DNA linear BCT 14-AUG-2003
DEFINITION Bordetella bronchiseptica strain RB50, complete genome; segment
15/16.
ACCESSION  BX640451 BX470250
VERSION     BX640451.1 GI:33577672
KEYWORDS    complete genome
SOURCE      Bordetella bronchiseptica RB50
ORGANISM    Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Bordetella.
REFERENCE   1
AUTHORS     Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.,
Harris D.E., Holden M.T.G., Churcher C.R., Bentley S.D.,
Mungall K.L., Cerdano-Farraga A.M., Temple L., James K., Harris B.,
Quail M.A., Achtman M., Atkin R., Baker S., Basham D., Bason N.,
Cherevach I., Chillingworth T., Collins M., Cronin A., Davis P.,
Doggert J., Feltwell T., Goble A., Hamlin N., Hauser H.,
Holroyd S., Jagels K., Leather S., Moule S., Norberczak H.,
O'Neill S., Ormond D., Price C., Rabinowitsch E., Rutter S.,
Sanders M., Saunders D., Seeger K., Sharp S., Simmonds M.,
Skellton J., Squares K., Squares S., Stevens K., Unwin L.,
Whithead S., Barrall B.G. and Maskell D.J.
TITLE       Comparative analysis of the genome sequences of Bordetella
pertussis, Bordetella parapertussis and Bordetella bronchiseptica
JOURNAL     Online Publication
REMARK      Nat. Genet. DOI:10.1038/Ngl227-10.1038/Ngl227
REFERENCE   2 (bases 1 to 343473)
AUTHORS     Sebahia M.
DIRECT SUBMISSION
TITLE       Submitted on behalf of the Pathogen
JOURNAL     Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: ms@sanger.ac.uk
FEATURES    Location/Qualifiers
             1..343473
             /organism="Bordetella bronchiseptica RB50"
             /mol_type="genomic DNA"
             /strain="RB50"

/locus_tag="taxon:257310"
290..1084
/locus_tag="BB4576"
290..1084
/locus_tag="BB4576"
/notes="ortholog of Bordetella pertussis (BX470248) BP3319"
/codon_start=1
/transl_table=11
/product="putative IclR-family transcriptional regulator"
/protein_id="CAE34938.1"
/db_xref="GI:33577673"
/translation="NRGIRHMDKTLKGLTVLEAIRSLPGATFTLQIAEEIGLTR
SNTHRTLOTHAGVRRDPATGSYQCTIKLFELGARQLASLDARKLAPPFNGELARL
NYTHVSLDGLDVVVDKLDSPQLLAYSGVGRAPAYAVATCKALLACOTAYLD
HYGAALKHTEDTHVSLAKDELADVNRIGVAINRGWRQIGIGIAAIFNGLDKAV
AALGISGFLRLTNQMRREYARLVQQAARQLSAHNGYRGYFGGA"
misc_feature 494..1045
/locus_tag="BB4576"
/notes="HMMPFam hit to PF01614, DE Bacterial
transcriptional regulator, score 1.3e-54"
1101..2261
/locus_tag="BB4577"
1101..2261
/locus_tag="BB4577"
/notes="ortholog of Bordetella pertussis (BX470248) BP3318"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAE34939.1"
/db_xref="GI:33577674"
/translation="NQKSTLAGVRVIEICNVAGPFCOMLLADMGADVVKLEPPGTG
DTRSWAPITDGYSENFLNKRKSVTLNKNKNEARDAIELIAGADVLIENRPGV
MDRLGQCGQARPNRLVYCSVSGSPRQSGDFDLTIQNSGLMSVTEEGGA
PVKCGVPCDFTAGLYAAFAIAALRRVQADQGTHTIDVSMGLGTLAALQTSYFG
TGADPVKLSGHPNAPQVFCNKNGYFGMAAGNNSLWKAVCKVVERDDLADPRFAN
PDRAEHVALRDILEAFEDSHAWLERPQAGVPCAPINDYSEVLADPOVHMCW
VDLELPNGVTRTFVSPVPDRTAGVWRPFPALGEHNDEILSALAARPA"
1311..1877
/locus_tag="BB4577"
/notes="HMMPFam hit to PF02515, DE CAIB/BAIF family, score
6.1e-59"
2258..3001
/locus_tag="BB4578"
2258..3001
/locus_tag="BB4578"
/notes="ortholog of Bordetella pertussis (BX470248) BP3317"
/codon_start=1
/transl_table=11
/product="putative enoyl-CoA hydratase/isomerase"
/protein_id="CAE34940.1"
/db_xref="GI:33577675"
/translation="MSAMSASLPLAERRPAWTTLSRPEKRNALSALVELALDGV
DAAREQVPLVFPAGAGNFSAGFDFTDYETSEGDLRLRMVRIEMLQRVAGSPSLT
LALAHGRNFGAGVDLFAACKWRYCTPEAGFRMFLGLTFRFRDITVGADQALSI
LGSRAFADQARRIGVRDCAQAQWPAIDAAAATAALDPATRLHRLVRDDHD
DADLAALASRAAPQGLKARIDYLAQPA"
2312..2806
/locus_tag="BB4578"
/notes="HMMPFam hit to PF00378, DE Enoyl-CoA
hydratase/isomerase family, score 7.7e-10"
complement(2998..3873)
/genes="copB"
/locus_tag="BB4579"
complement(2998..3873)
/genes="copB"
/locus_tag="BB4579"
/locus_tag="BB4579"
/notes="ortholog of Bordetella pertussis (BX470248) BP3316"
/codon_start=1
/transl_table=11
/product="copper resistance protein"
/protein_id="CAE34941.1"
/db_xref="GI:33577676"
/translation="NAASLGLAAAAAASAWAQAPAMDHGSMGMDHGAAPASAMDHGSM
```



misc\_feature  
 PCMDGGAAPAPALPSSDGDGDSYAAVGIHSHMDDMATSSOLLFDKLGTVDRDR  
 RTGLWQDFWVGRDLKXWIKSESERVGGSDGKVEAFWNRITISPFWDLOVGARRDR  
 GTGKRNKAGLQGVAVNIETETVATVGGSGRTALAKAEYDILLITQRLVLPETIE  
 ASVYGDKARGIGAGLSDFSLSRLYEITREVPVYGVGSGRKFORTASYASEGE  
 SRSRAINAVGRIWTF  
 complement(3826..3873)  
 /gene="copA"  
 /locus\_tag="BB4579"  
 /note="Signal peptide predicted for BB4579 by SignalP 2.0  
 HMM (Signal peptide probability 1.000) with cleavage site  
 probability 0.953 between residues 16 and 17;  
 signal-peptide site"  
 complement(3920..5677)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /locus\_tag="BB4580"  
 complement(3920..5677)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="ortholog of Bordetella pertussis (BX470248) BP3315"  
 /codon\_start=1  
 /transl\_table=11  
 /product="copper resistance protein"  
 /protein\_id="CAE34942.1"  
 /db\_xref="GI:33577677"  
 /translation="MSRAVSNQRRFFVGLAGGVLAGLWRAQAAWPGAAQPEVL  
 SOTENLEIGTTPANFGAARMAATGYSIPAPILRWREGDTVLRVTNLDSDTSIH  
 WHGILLPTGMDGVPGLRPGFTFYRFQSGSTWYVYHSHSGFQEGTGLYGAIV  
 TPRADPVRAERDHWLLSDWTDPMVSKLVKMPDYNYIQPTDNLVDADARKO  
 CHASALGERLWQWQNMOTDLADYSGATVYLANGKSPAGNWTGLPKGERVRLRI  
 NGSAMTYDVRILPGLNTVAADGLVPAPVEDEFRIVASTYDVIIVEPADRAATYIF  
 SOSMDSRGVRAIAPRGSGHAEVPALDPVQMLTMDMGMAHDPVPEHGAAMDHGM  
 GMDHAAHGHSGEASGVREHPYPAEHGPAANSMLPDIVITRLDPPGVGLRDGRR  
 VLTYADLRSMVEPNRAPTRIEHLTGNDYRYSFNGLKFQAKPIVLKYGERVR  
 FVLVNDTWMTPIHLHGWSDLSPDGNFQVRKHITSLNPAQRIYSRVSDARGNWAY  
 HCHLLYHMEAGMFRAVVVE"  
 complement(5576..5677)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="Signal peptide predicted for BB4580 by SignalP 2.0  
 HMM (Signal peptide probability 0.962) with cleavage site  
 probability 0.882 between residues 34 and 35;  
 signal-peptide site"  
 complement(3926..4324)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="this hit extended beyond the end of the feature by  
 1 aa and was clipped.  
 HMPpfam hit to PF00394, DE Multicopper oxidase, score  
 0.0021"  
 complement(3932..3994)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="PS00079 Multicopper oxidases signature 1."  
 complement(3974..3979)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="PS00080 Multicopper oxidases signature 2."  
 complement(4619..5176)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="HMPpfam hit to PF00394, DE Multicopper oxidase,  
 score 1.2e-17"  
 complement(4850..4879)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="PS00215 Mitochondrial energy transfer proteins  
 signature."  
 complement(5183..5557)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="HMPpfam hit to PF00394, DE Multicopper oxidase,  
 score 3.2e-08"  
 complement(5201..5263)  
 /gene="copA"

misc\_feature  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="PS00079 Multicopper oxidases signature 1."  
 complement(5576..5644)  
 /gene="copA"  
 /locus\_tag="BB4580"  
 /note="1 probable transmembrane helix predicted for BB4580  
 by TMHMM2.0 at aa 12-34"  
 complement(5721..6215)  
 /locus\_tag="BB4581"  
 complement(5721..6215)  
 /locus\_tag="BB4581"  
 /note="ortholog of Bordetella pertussis (BX470248) BP3314"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative blue copper protein"  
 /protein\_id="CAE34943.1"  
 Query Match 91.6%; Score 17.4; DB 1; Length 343473;  
 Best Local Similarity 94.7%; Pred. No. 3.2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GTCGAGCGCCGCGATGACC 19  
 Db 68738 GCCGACGCCGCGATGACC 68756  
 RESULT 17  
 BX640435  
 LOCUS BX640435 346259 bp DNA linear BCT 14-AUG-2003  
 DEFINITION Bordetella parapertussis strain 12822, complete genome; segment  
 13/14.  
 ACCESSION BX640435 BX470249  
 VERSION BX640435.1 GI:33574489  
 KEYWORDS complete genome.  
 SOURCE Bordetella parapertussis  
 ORGANISM Bordetella parapertussis  
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 Alcaligenaceae; Bordetella.  
 REFERENCE 1  
 AUTHORS Parkhill, J., Sebaihia, M., Preston, A., Murphy, L. D., Thomson, N.,  
 Harris, D. E., Holden, M. T. G., Churcher, C. R., Bentley, S. D.,  
 Mungall, K. L., Cerdano-Tarraga, A. M., Temple, L., James, K., Harris, B.,  
 Chuai, M. A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N.,  
 Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P.,  
 Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H.,  
 Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H.,  
 O'Neill, S., Ormond, D., Price, C., Rabinowitsch, E., Rutter, S.,  
 Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M.,  
 Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L.,  
 Whitehead, S., Barrell, B. G. and Maskell, D. J.  
 Comparative analysis of the genome sequences of Bordetella  
 pertussis, Bordetella parapertussis and Bordetella bronchiseptica  
 Online Publication  
 Nat. Genet. DOI:10.1038/Ng1227-10.1038/Ng1227  
 REFERENCE 2 (bases 1 to 346259)  
 AUTHORS Sebaihia, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-2003) Submitted on behalf of the Pathocen  
 Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA, E-mail: ms5@sanger.ac.uk  
 FEATURES  
 source  
 1. 346259  
 /organism="Bordetella parapertussis"  
 /mol\_type="genomic DNA"  
 /strain="12822"  
 /db\_xref="taxon:519"  
 complement(293..937)  
 /gene="ureG"  
 /locus\_tag="BPP3852"  
 complement(293..937)  
 /gene="ureG"  
 /locus\_tag="BPP3852"



```

/note="ortholog of Bordetella pertussis (BX470248) BP3166"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39135.1"
/db_xref="GI:33574490"
/translation="MHDISSLTTRTKTLPLRVLGVGPGVSGKTKTLLVMCKWAMYPOF
DLIAITNDIVYKEQRLTLTSGALPPERILGVETGCGPHTAIRSDASINLIADQMLE
QPADIVFVSEGDNLAAATFSPSLDLTIYIIDVSGEKIPKGGGIGTKSLDFIIN
KTDLAPVVGADLAWEADTRMRGDKPFVNCNLKTGDLQDVIAFLKTEGLFRG"
complement(509..892)
/misc_feature
/gene="ureG"
/locus_tag="BPP3852"
/note="HMPFfam hit to PF01495, HypB/UreG
nucleotide-binding domain, score 4.5e-73"
complement(848..871)
/misc_feature
/gene="ureG"
/locus_tag="BPP3852"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)."
complement(962..1624)
/gene="ureF"
/locus_tag="BPP3853"
complement(962..1624)
/gene="ureF"
/locus_tag="BPP3853"
/note="ortholog of Bordetella pertussis (BX470248) BP3167"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39136.1"
/db_xref="GI:33574491"
/translation="MHLSSPALPIGFGFSYSGLEAAETELGLVHDEASTLAWIRSQLV
VMAEAPLCLLFEANRAGDDAAAGNQHFRASRETRILROETEQMGSLRLAQE
LGNQTAALRAATLPAFVHACAMWALPREAGLGAYFVSWLENOVAAAIKGV
PLGQVAGQRMLEURLAGLAVLADARARAGATPRLDTFAPQVALVSARHETQFSRL
RS"
complement(1094..1546)
/misc_feature
/gene="ureF"
/locus_tag="BPP3853"
/note="HMPFfam hit to PF01730, UreF, score 9.7e-74"
complement(1647..2264)
/gene="ureB"
/locus_tag="BPP3854"
complement(1647..2264)
/gene="ureB"
/locus_tag="BPP3854"
/note="ortholog of Bordetella pertussis (BX470248)
BP3167A"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39137.1"
/db_xref="GI:33574492"
/translation="MKIANFIKGAAGGLDLSQAPGVTLTAERRSRORLDLDEGR
GELGMATERTGLRDGDLVVAEDGTYYVVRRAALDVARVTAATPWQLARAYHLGNRH
VLLTAQHLQFEDAVLIDMLAQLGVTAMRLRAVFPDVGAYGGHGHDESFQD
DYALQAAYHAHEAPHASHAGGHGVHSHGHGKHGHEAAS"
complement(1944..2249)
/gene="ureB"
/locus_tag="BPP3854"
/note="HMPFfam hit to PF02814, UreE urease accessory
protein, score 1.7e-45"
complement(2267..3982)
/gene="ureC"
/locus_tag="BPP3855"
complement(2267..3982)
/gene="ureC"
/locus_tag="BPP3855"
/EC_number="3.5.1.5"
/note="ortholog of Bordetella pertussis (BX470248) BP3168"
/codon_start=1
/transl_table=11
/product="urease alpha subunit"

```

```

/protein_id="CAE39138.1"
/db_xref="GI:33574493"
/translation="MTRISRSAYAIYGVTVVGGVDRVRLADTILLAEVEKDHITFG
EVKFGGKVIDGMGQRLATCDVDTVITNALLIDAVTGIVKADIGIKGLGISGIG
KAGNPTQGVITIIIGASTEVVAGELIVTAGAIDTHIFICFOQIEBALATGTTMI
GGTGPATGSLATTISGFWMAAMQLQALDAPFVNVGLFGKSSSSHGALLBQVRAGA
MGLKIHEDWASTPASIDICLVNAEETDIOVAIHSDTLNESGFVEDTFAAFKGRTHSF
HTEGAGGHAPDIIRAAGPNVLPASTNPTMFTTRTIDHLDMMVMVCHLDPISAEED
LAFASRRRTTIAEEDILHDIGAFSINMSDSQAMGRVGEIVLRTWQAHKMKLQKGP
LOQDSRVSERIKRYVITNPVAVAGIAHLVGSVEVGLADLVLPKPAFFGVKVN
MYLKSGMAVSASIKGMGASISTPOPQVIRPMKSHGKALRTSVAFVSOVSLSPAVSE
LELNKLEAVRCRGVTKHDMVNNWLPASIVDPQTYQVYADGQLLRCEALAEPLMAQ
RYLFL"
complement(2273..3580)
/misc_feature
/gene="ureC"
/locus_tag="BPP3855"
/note="HMPFfam hit to PF02802, Urease alpha-subunit,
catalytic domain, score 0"
complement(2969..3019)
/misc_feature
/gene="ureC"
/locus_tag="BPP3855"
/note="PS00145 Urease active site."
complement(3548..3589)
/misc_feature
/gene="ureC"
/locus_tag="BPP3855"
/note="PS01120 Urease nickel ligands signature."
complement(3584..3979)
/misc_feature
/gene="ureC"
/locus_tag="BPP3855"
/note="HMPFfam hit to PF00449, Urease alpha-subunit,
N-terminal domain, score 4.8e-83"
complement(3979..4287)
/gene="ureB"
/locus_tag="BPP3856"
complement(3979..4287)
/gene="ureB"
/locus_tag="BPP3856"
/EC_number="3.5.1.5"
/note="ortholog of Bordetella pertussis (BX470248) BP3169"
/codon_start=1
/transl_table=11
/product="urease beta subunit"
/protein_id="CAE39139.1"
/db_xref="GI:33574494"
/translation="MIPGELTPQGIELNVGRPTLTIAVNEGRPIQVGSYHFAE
ANALVDFRELATGYRLNPAGNAVPEFGMERTVELVAVGGERRIFGQGVMGALK"
complement(3985..4284)
/misc_feature
/gene="ureB"
/locus_tag="BPP3856"
/note="This hit extended beyond the end of the feature by
1 aa and was clipped.
HMPFfam hit to PF00699, Urease beta subunit, score
3.2e-60"
complement(4304..4921)
/gene="ureJ"
/locus_tag="BPP3857"
complement(4304..4921)
/gene="ureJ"
/locus_tag="BPP3857"
/note="ortholog of Bordetella pertussis (BX470248) BP3170"
/codon_start=1
/transl_table=11
/product="urease accessory protein"
/protein_id="CAE39140.1"
/db_xref="GI:33574495"
/translation="MKAMSKRAMLGSGRAALMLFSGAALAHPHGLGHELPGSMFAAGF
WHLPGFDHLLAVLAVGMWSALHTHTARQAVLWPNVFLALLPAGMMGAGVRLPAVE
PVMVSLVILGLLVASRKAVQWAGFALVGFALPHGLAHGMEIPLGSEGALGFVAGFM
LATLGLHLGLFAGFLKFKLKNLWLSRALGVGTAGYCALLFVGARY"
complement(4844..4921)
/misc_feature
/gene="ureJ"
/locus_tag="BPP3857"

```

```

/note="Signal peptide predicted for BPP3857 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 1.000 between residues 26 and 27;
signal-peptide site"
complement(join(4310..4375,4397..4462,4493..4549,
4571..4624,4637..4702,4724..4789,4835..4900))
/genes="ured"
/locus_tag="BPP3857"
/note="7 probable transmembrane helices predicted for
BPP3857 by TMHMM2.0 at aa 7-29, 44-66, 73-95, 100-117,
91.6%; Score 17.4; DB 1; Length 346259;
Best Local Similarity 94.7%; Pred. No. 3.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

misc_feature
1 GTCGAGCGCGCGATGACC 19
|||||
Db 333562 GCGAGCGCGCGATGACC 333580

RESULT 18
EX640412/c
LOCUS
DEFINITION
BX640412 348171 bp DNA linear BCT 14-AUG-2003
Bordetella pertussis strain Tohama I, complete genome; segment
2/12
ACCESSION
EX640412 BX470248
VERSION
EX640412.1 GI:33571216
KEYWORDS
complete genome.
SOURCE
Bordetella pertussis Tohama I
Bordetella pertussis Tohama I
ORGANISM
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Bordetella.
REFERENCE
1 Parkhill,J., Sebahia,M., Preston,A., Murphy,L.D., Thomson,N.,
Harris,D.E., Holden,M.T.G., Churcher,C.R., Bentley,S.D.,
Mungall,K.L., Cerdono-Tarraga,A.M., Temple,L., James,K., Harris,B.,
Quail,M.A., Achtman,M., Atkin,R., Baker,S., Basham,D., Bason,N.,
Cherevach,I., Chillingworth,T., Collins,M., Cronin,A., Davis,P.,
Duggett,J., Feltwell,T., Goble,A., Hamlin,N., Hauser,H.,
Holroyd,S., Jagers,K., Leather,S., Moule,S., Norberczak,H.,
O'Neill,S., Ormond,D., Price,C., Rabinowitch,B., Rutter,S.,
Sanders,M., Saunders,D., Seeger,K., Sharp,S., Simmonds,M.,
Skellton,J., Squares,R., Squares,S., Stevens,K., Unwin,L.,
Whitehead,S., Barrall,B.G. and Maskell,D.J.
Comparative analysis of the genome sequences of Bordetella
pertussis, Bordetella parapertussis and Bordetella bronchiseptica
Online Publication
Nat. Genet. DOI:10.1038/Ngl227-10.1038/Ngl227
2 (bases 1 to 348171)
Sebahia,M.
Direct Submission
Submitted (06-AUG-2003) Submitted on behalf of the Pathogen
Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, E-mail: mss@sanger.ac.uk
Location/Qualifiers
1. 348171
/organism="Bordetella pertussis Tohama I"
/mol_type="genomic DNA"
/strain="Tohama I"
/db_xref="taxon:257313"
complement(52..921)
/genes="bhuT"
/locus_tag="BP0345"
complement(52..921)
/genes="bhuT"
/locus_tag="BP0345"
/note="Similar to Yersinia enterocolitica hemin binding
protein HmT TR:P74979 (EMBL:X77867) (279 aa) fasta
scores: E(): 8.2e-29, 37.638% id in 271 aa, and to
Yersinia pestis hemin-binding periplasmic protein HmT
precursor SW:HMUT_YERPE (Q56591) (279 aa) fasta scores:
E(): 9.4e-29, 37.638% id in 271 aa"
/codon_start=1

FEATURES
source
/transl_table=11
/product="putative hemin binding protein"
/protein_id="CAE44677.1"
/db_xref="GI:33571217"
/transl_table=11
/note="Signal peptide predicted for BP0345 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.617 between residues 23 and 24;
signal-peptide site"
complement(148..840)
/genes="bhuT"
/locus_tag="BP0345"
/note="HMPfam hit to PF01497, Periplasmic binding
protein"
complement(835..900)
/genes="bhuT"
/locus_tag="BP0345"
/note="1 probable transmembrane helix predicted for BP0345
by TMHMM2.0 at aa 7-23"
complement(918..1964)
/genes="bhuS"
/locus_tag="BP0346"
complement(918..1964)
/genes="bhuS"
/locus_tag="BP0346"
/note="Similar to Yersinia pestis hemin transport protein
HmS SW:HMUS_YERPE (Q56590) (345 aa) fasta scores: E():
1.1e-46, 42.486% id in 346 aa, and to Yersinia
enterocolitica hemin transport protein HmS SW:HEMS_YEREN
(P31517) (345 aa) fasta scores: E(): 1.2e-46, 43.413% id
in 334 aa"
/codon_start=1
/transl_table=11
/product="putative hemin transport protein"
/protein_id="CAE44678.1"
/db_xref="GI:33571218"
/transl_table=11
/note="Signal peptide predicted for BP0347 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.617 between residues 23 and 24;
signal-peptide site"
complement(148..840)
/genes="bhuT"
/locus_tag="BP0347"
complement(148..840)
/genes="bhuT"
/locus_tag="BP0347"
/note="Similar to Yersinia enterocolitica hemin binding
protein HmT TR:P74979 (EMBL:X77867) (279 aa) fasta
scores: E(): 8.2e-29, 37.638% id in 271 aa, and to
Yersinia pestis hemin-binding periplasmic protein HmT
precursor SW:HMUT_YERPE (Q56591) (279 aa) fasta scores:
E(): 9.4e-29, 37.638% id in 271 aa"
/codon_start=1

```

```

/transl_table=11
/product="putative hemin binding protein"
/protein_id="CAE44677.1"
/db_xref="GI:33571217"
/transl_table=11
/note="Signal peptide predicted for BP0345 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.617 between residues 23 and 24;
signal-peptide site"
complement(148..840)
/genes="bhuT"
/locus_tag="BP0345"
/note="HMPfam hit to PF01497, Periplasmic binding
protein"
complement(835..900)
/genes="bhuT"
/locus_tag="BP0345"
/note="1 probable transmembrane helix predicted for BP0345
by TMHMM2.0 at aa 7-23"
complement(918..1964)
/genes="bhuS"
/locus_tag="BP0346"
complement(918..1964)
/genes="bhuS"
/locus_tag="BP0346"
/note="Similar to Yersinia pestis hemin transport protein
HmS SW:HMUS_YERPE (Q56590) (345 aa) fasta scores: E():
1.1e-46, 42.486% id in 346 aa, and to Yersinia
enterocolitica hemin transport protein HmS SW:HEMS_YEREN
(P31517) (345 aa) fasta scores: E(): 1.2e-46, 43.413% id
in 334 aa"
/codon_start=1
/transl_table=11
/product="putative hemin transport protein"
/protein_id="CAE44678.1"
/db_xref="GI:33571218"
/transl_table=11
/note="Signal peptide predicted for BP0347 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.617 between residues 23 and 24;
signal-peptide site"
complement(148..840)
/genes="bhuT"
/locus_tag="BP0347"
complement(148..840)
/genes="bhuT"
/locus_tag="BP0347"
/note="Similar to Yersinia enterocolitica hemin binding
protein HmT TR:P74979 (EMBL:X77867) (279 aa) fasta
scores: E(): 8.2e-29, 37.638% id in 271 aa, and to
Yersinia pestis hemin-binding periplasmic protein HmT
precursor SW:HMUT_YERPE (Q56591) (279 aa) fasta scores:
E(): 9.4e-29, 37.638% id in 271 aa"
/codon_start=1

```

```
TSWLLQAGTRNGHLDNRADTGYGSKRSQSPEDYAOANFLKLQORIDGHRGLGT
GEYFRADLDOMYQOCAGTSYGYGANRTHETTRKRVSLDYQYNAPQAGAADISARA
MYVQRLRLSSODARETRDGRAVARPDYFYGYPSGPGYGRNSIOESILGVNGELS
PRFGYVSQVITIGEGYNGARTESYSGYDNCNPAIPGTPAPGPRICLMLNGLQADM
PRVGSQMAWAQDEIADPAGRIYLTSLRYDHYEQPQGGQNNPNAGALPPSSS
GREFPKLLGRLGALTYAOCYFGYRAPSATELTYNGPCTYLVRGNPLKPEET
SKWELGARLGDOLGAVSLFDNRYQNFIDKNVPLGKSPQWQMDGQYPLGVITGL
ANRWIYGAESAARHWPNTWGLAWGKDNCTGHLNSVPLKAILGLGYQ
RDEWIGIDAMLTAAVTRDDVOYPEASARVADFOAGYGVVDLSAYWPAVAVKGLQLQ
AGVNLFDKKWEALNVPITAGAIAPRLDWNPEGRSVRSVLTYYQY
/locus_tag="BP0347"
/note="Signal peptide predicted for BP0347 by SignalP 2.0
HMM (Signal peptide probability 1.000) with cleavage site
probability 0.975 between residues 21 and 22;
signal-peptide site"
/complement(2003..2371)
/gene="bhuk"
/locus_tag="BP0347"
/note="HMPfam hit to PF00593, TonB dependent receptor
C-terminal region"
/complement(4745..5662)
/gene="huk2"
/locus_tag="BP0348"
/complement(4745..5662)
/gene="huk2"
/locus_tag="BP0348"
/note="Similar to Escherichia coli protein FecR or B4292
SW:FECD ECOLI (P23485) (317 aa) fasta scores: E(): 1e-22,
32.459% id in 305 aa, and to Pseudomonas aeruginosa
probable transmembrane sensor Pa911 TR:Q912J3
(EMBL:AE004617) (316 aa) fasta scores: E(): 2.4e-32,
41.118% id in 304 aa"
/codon_start=1
/transl_table=1
/product="heme uptake transmembrane sensor"
/protein_id="CAE44680.1"
/db_xref="GI:33571220"
/translation="MADEAARWLVLGSGQASAEIQCDHWRASHAEHRAWCRARR
LTSFDRIPVAGQALGRADRRAMLKSLVALLAAPAAALRGARNSGWLADLRT
GTGRTVVALPGTQLRLTGTVDLSDGLSLRLHGEIYLEADRPCRVLTTRGM
IRTOAHLWLQDDADGLGVVAGLAFWQAGDNTHSVAAGRIAWHDGALQGAOITL
DGSQDLWGLVLRADAMLDRLRLSRYRPTCLCDPRVAGLRSLGVFLAHTDILR
ALPALLFQLSIVTPIYVITVGFPRAGAIT"
/complement(5606..5662)
/gene="huk2"
/locus_tag="BP0348"
/note="Signal peptide predicted for BP0348 by SignalP 2.0
HMM (Signal peptide probability 0.708) with cleavage site
probability 0.708 between residues 19 and 20;
signal-peptide site"
/complement(5698..6207)
/gene="huri"
/locus_tag="BP0349"
/complement(5698..6207)
/gene="huri"
/locus_tag="BP0349"
/note="Similar to Escherichia coli probable RNA polymerase
sigma factor FecI or B4293 SW:FEI ECOLI (P23484) (173 aa)
fasta scores: E(): 1.2e-26, 50.649% id in 154 aa, and to
Pseudomonas aeruginosa putative RNA polymerase sigma
factor Pa1912 TR:Q912J2 (EMBL:AE004617) (168 aa) fasta
scores: E(): 8.9e-31, 55.699% id in 167 aa. Also similar
to Bpl136, 46.875% identity in 160 aa overlap"
/codon_start=1
/transl_table=1
/product="heme uptake regulator"
/protein_id="CAE44681.1"
/db_xref="GI:33571221"
/translation="MPLTSLAPQDAVEALYREHGWLQSLKLNLDAYDAADLAQDT
FVRLVRAEATAREFRAYLTITAGRLVNVHRRSLRAYLEALSLPAPAPSE
QRALILSALDQIDAFLLAPKPVQAFLLAQLERLPYADIARRLQVSVRTVQRIAG
```

```
YEOCILLAP"
misc_feature complement(5734..5799)
/gene="huri"
Query Match 91.6%; Score 17.4; DB 1; Length 348171;
Best Local Similarity 94.7%; Pred. No. 3.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GTCGACGCCGCCGATGACC 19
Db 15873 GCGGACGCCGCCGATGACC 15855
RESULT 19
ADP279815/c 7336 bp DNA linear VRL 15-MAY-2001
LOCUS Ascovirus DPAV4 PstI genomic DNA fragment, ORF1-ORF15.
DEFINITION AJ279815
ACCESSION AJ279815
VERSION GI:14031079
KEYWORDS ORF1; ORF10; ORF11; ORF12; ORF13; ORF14; ORF15; ORF2; ORF3; ORF4;
ORF5; ORF6; ORF7; ORF8; ORF9.
SOURCE Ascovirus DPAV4
ORGANISM Ascovirus DPAV4
Viruses; dsDNA viruses, no RNA stage; Ascoviridae; Ascovirus.
REFERENCE 1
AUTHORS Stasiak,K., Demattei,M.V., Federici,B.A. and Bigot,Y.
TITLE Phylogenetic position of the Diadromus pulchellus ascovirus DNA
JOURNAL polymerase among viruses with large double-stranded DNA genomes
MEDLINE J. Gen. Virol. 81 (Pt 12), 3059-3072 (2000)
PUBMED 20540044
11086137
REFERENCE 2
AUTHORS Bigot, Y.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2000) Bigot Y., IRBI, Groupe d'etude des
Parasites moleculaires, Faculte des sciences, Parc Grandmont, Avenue
Monge, 37200 Tours, FRANCE
REMARK 3 (bases 1 to 6736)
AUTHORS Bigot, Y.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-2001) Bigot Y., IRBI, Groupe d'etude des
Parasites moleculaires, Faculte des sciences, Parc Grandmont, Avenue
Monge, 37200 Tours, FRANCE
COMMENT On May 14, 2001 this sequence version replaced gi:11931735.
Related sequences: ORF3 - Q00196 (SP), ORF2 P14358 (SP), ORF1 -
O55720 (SP), ORF5 - O55733 (SP), ORF6 - O55763 (SP), ORF7 -
CEF19B2 2.
FEATURES Location/Qualifiers
1..6736
/organism="Ascovirus DPAV4"
/virus
/mol_type="genomic DNA"
/specific_host="Diadromus pulchellus"
/db_xref="taxon:113365"
/lab_host="Diadromus pulchellus"
complement(<1..990)
/note="ORF2"
Fv3 Iridoviral protein homologue"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19145.1"
/db_xref="GI:11931737"
/db_xref="SPTREMBL:Q9DSU0"
/translation="VAIITKSLIERTEDAADHGIRITDELGLNFCADMPPPTDL
ERKRGVFDGKLVFRSSFFPTGLPESLDGFKVFTKMGVIVIRFFNTKMYVSSN
RKLDAPKSKWGAISFGEIFFNTVEKLKGSDEFLESDKSGIYLFRMGTIPTSTIVA
PAGYDLEILNSFDMSGSEKIDENSLSYNTKQFSLNKDVEEYVENMQMPFSDGYG
FLYNPDAPCKLENSTYKKLSRLHNHVSIPPAYLHNCFDENNKKFRDLYPHSETFD
FYDSEINAI SANLLNVIYIKIHVDKQKIEVSKQEHVLYVHVGKLYSSRQPTTLQ"
560..1039
/note="ORF3"
murine protein homologue"
```

```

/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19146.1"
/db_xref="GI:11931738"
/db_xref="SPTREMBL:Q9DST9"
/translation="MLSSDSKNSGDFPNSFVTKNISPKEVAPHLDLKASSFLLE
LTHVLAKKTLITVPSFMVTKLPSRLSGSGGKGGDGLNTLNLSPSNTTPTTLRSRS
VVGMSAQKPRSPSSVILMPWSAASSVRSINALFVLIAMFWVNLITSEFFNFH"
complement(1722..2318)
/notes="ORF4"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19147.1"
/db_xref="GI:11931739"
/db_xref="SPTREMBL:Q9DST8"
/translation="MNGKVSRCALDDARPSIKTSVCRLASAVLLESWMLNTEVES
WSMSRSLTUPPVFLTKANVVFVFLHSMDFPFVITRLIMLWLTFLVLAASCAP
SPSTRMADCSFSPSTLLVLSAFNKLPTLTLTRCVKVTLEMLTSPVPSFTTM
DWLLSEYAFICFVLATSVTAFTILVELVDIADPIYII"
1735..3060
/notes="ORF5"
CIV iridoviral protein homologue"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19148.1"
/db_xref="GI:11931740"
/db_xref="SPTREMBL:Q9DST7"
/translation="WGSALSTNSTFLVMKAVTEVANTNIOKAYSDDNQSIVVKIENG
GDVNNITFORVNNVGNLNTNNEVLNKLQKASMLVREGNLNLAQAAST
NVNVLINLVTKNRSINECKSTTKTFSLOVRNTGNNVNRDLIDQDSSTVFKCI
QDSKSTADLRQTDVLDGLASSAQAQGLDRLVIGGVGVALIGVGOTFIGAKVIGP
VIASPELLASTDSTNSPRTQLIKTNKVKSLKNATPTFTPAQILSKYKSAEN
LTDIIRGLQADVLEGTVLYKSAKPRDAYDADDGLGPVALIIDVDNTRGI
SVKLONLGHNFANFVNLINLKEDMKAAGDVAGQYDOSTHTVSVFKEDDPTFTSN
VPASVVMIKPEKPKPVNKQLVYIGYALLATGWTATMSFRNGSGGTRS"
3171..3569
/notes="ORF6"
CIV iridoviral protein homologue"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19149.1"
/db_xref="GI:11931741"
/db_xref="SPTREMBL:Q9DST6"
/translation="MNPYSAPKLTDELRIIGDKEVQIVIVERQWQSHAIYRVASHT
ENYFAMVPSAEYIVCDPKLDLKGVGKGAARKTAKCAKKYQOEGLDLASIDA
HKKDDMDCAICQWLWFCANRVKFNVTIK"
4004..4315
/notes="ORF7"
nematod protein homologue"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19150.1"
/db_xref="GI:11931742"
/db_xref="SPTREMBL:Q9DST5"
/translation="MLKNVTVICIFRALSSFWLSIWSHFFLWVICLAAEMGLON
LITTSRRFSLKLYLLSMGLTMSLTVIPGSLNMGALGDPVFRNCKSVLFGVDNALIL
V"
complement(4301..4664)
/notes="ORF8"
CDS
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19151.1"
/db_xref="GI:14031081"
/db_xref="SPTREMBL:Q9DST4"
/translation="MRRPRYFRSNFTFOINNFRGEVNNVVARLIHETTFENFVSIL
KSKMLYKPSNNLNGTVQGSRRNRIVSDPFISLKNKNFYDYDEVDAVYMLYDEDTF
VHDDVMFIFDPDSSIPV"
complement(4302..4664)
/notes="ORF9"
CDS
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19152.1"
/db_xref="GI:11931743"
/db_xref="SPTREMBL:Q9DST3"
/translation="MRRPRYFRSNFTFOINNFRGEVNNVVARLIHETTFENFVSIL
KSKMLYKPSNNLNGTVQGSRRNRIVSDPFISLKNKNFYDYDEVDAVYMLYDEDTF
VHDDVMFIFDPDSSIPV"
complement(4670..5072)
/notes="ORF14"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19153.1"
/db_xref="GI:14031082"
/db_xref="SPTREMBL:Q9DST3"
/translation="MEWKIDAWKFNCRVLLFRFTDAFKKASVDSIRVLVQKTL
QRTVDLEAPITDVMNMLFETPAATVLAVALFEATQALDSKRRQKIYSVVGVED
LRDVMSTLSEDKTMPSORHVEGDRFNQV"
complement(4671..5072)
/notes="ORF10"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19154.1"
/db_xref="GI:11931744"
/db_xref="SPTREMBL:Q9DST3"
/translation="MEWKIDAWKFNCRVLLFRFTDAFKKASVDSIRVLVQKTL
QRTVDLEAPITDVMNMLFETPAATVLAVALFEATQALDSKRRQKIYSVVGVED
LRDVMSTLSEDKTMPSORHVEGDRFNQV"
4888..5157
/notes="ORF11"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19155.1"
/db_xref="GI:11931745"
/db_xref="SPTREMBL:Q9DST2"
/translation="MFQSTVICASRSTVRCRVFWRITLMESTDAFFLKASVNSLNN
NTRLQNFHAFIFHSMKNPSETLDSRSRSVTPRSLMGRTHL"
complement(5437..56047)
/notes="ORF1"
CIV iridoviral protein homologue"
/codon_start=3
/product="hypothetical protein"
/protein_id="CAC19156.1"
/db_xref="GI:11931736"
/db_xref="SPTREMBL:Q9DST1"
/translation="AGRTSTRHAGGRQRPHELFDRLQIQRORRQOLYARKTETR
DHSILVRGVLISQUPQTNLHHEHKGQGGRRSVLEGRHQRNGKRPYFRIPGQ
FNTSGIYDASGGGSGFRISTGRSRRARRARFVNDLHTSNVVKCKDNKLFIL
YRPFAGKRVKLVTRGYPIVYIDYLSYSSECDGMSRDL"
complement(join(<5440..6079,6404..6513))
/notes="unnamed protein product; ORF15"
/codon_start=1
/protein_id="CAC39841.1"
/db_xref="GI:14132804"
/db_xref="SPTREMBL:Q9IU95"
/translation="MPSSAIHPALRPLQNCILLIAEGTLVLLQSLTSLKSTGFRV
TLEATGTRTHAGGRQRPHELFDRLQIQRORRQOLYARKTETRDSHSLRGVL
SLSQLPOTNLHHEHKGQGGRRSVLEGRHQRNGKRPYFRIPGQFNTSGIYD
ASGGSGFRISTGRSRRARRARFVNDLHTSNVVKCKDNKLFILYRFDFAKGR
VALVTRGYPIVYIDYLSYSSECDGMSRDL"
complement(5440..6079)
/notes="ORF15"
/number=2
complement(5523..5849)
/notes="ORF12"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAC19154.1"
/db_xref="GI:11931746"
/db_xref="SPTREMBL:Q9DST1"
/translation="MKNIKVAKDGASPKVATDKTVNVDLIGFIEYLDNFVTLHKF
IRTHPAEVEHVSFVQLVIALVVRAGALVSCITCTLTPLWLSKNVIKYSYCTASTSRV

```

Query Match 89.5%; Score 17; DB 14; Length 6736;  
 Best Local Similarity 100.0%; Fred. No. 6.3e+03; Indels 0; Gaps 0;  
 Matches 17; Conservative 0; Mismatches 0;

```
QY      3  CGACGCGCGCGATGACC 19
Db      2329 CGACGCGCGCGATGACC 2313
|||||
|||||

RESULT 20
AB005103
LOCUS      11252 bp      DNA      linear      BCT 12-FEB-2001
DEFINITION Halobacterium sp. NRC-1 section 134 of 170 of the complete genome.
ACCESSION AB005103 AB004437
VERSION    AB005103.1 GI:10581562
KEYWORDS
SOURCE
ORGANISM  Halobacterium sp. NRC-1
            Halobacterium sp. NRC-1
            Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
            Halobacteriaceae; Halobacterium.
REFERENCE 1 (bases 1 to 11252)
AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Laaky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
            Genome sequence of Halobacterium species NRC-1
            Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
            20504483
            11016950
REFERENCE 2 (bases 1 to 11252)
AUTHORS   Ng,W.V., Kennedy,S.P., Mahairas,G.G., Berquist,B., Pan,M.,
            Shukla,H.D., Laaky,S.R., Baliga,N., Thorsson,V., Sbrogna,J.,
            Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,
            Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
            Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
            Isenbarger,T.A., Peck,R.F., Pohlschrod,M., Spudich,J.L.,
            Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
            Omer,A.D., Ebhardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
            DasSarma,S.
            Direct Submission
            Submitted (14-JUL-2000) Institute for Systems Biology, 4225
            Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES
source
1..11252
    /organism="Halobacterium sp. NRC-1"
    /mol_type="genomic DNA"
    /strain="NRC-1"
    /db_xref="taxon:64091"
68..724
    /gene="hmp"
    /note="synonym: VNG2147G"
68..724
    /gene="hmp"
    /note="hmp"
    /codon_start=1
    /transl_table=11
    /product="membrane protein"
    /protein_id="AAG20284.1"
    /db_xref="GI:10581563"
    /translation="MRVQANRSRGVLEDKARTFYKYLKVDVNTFLWTEWRA
EALSMLDIAEDRVLDVGCCTGCTGTEGLKXTHVYGLDQSPHQLKAEKFGKHPV
AFHFGDLPFPKDSIEYWPHPDQGLRUVAKPQSGVLVVGPDYPA
STVFRKLADAILMFYDEDEADRMFRFAGYEIEIQRVLRKEPNPFAIVTVARAADD
"
    complement (1227..2462)
    /gene="VNG2148H"
    complement (1227..2462)
    /gene="VNG2148H"
    /note="hypothetical protein"
    /codon_start=1
    /transl_table=11

gene
CDS
    complement (1227..2462)
    /gene="VNG2148H"
    /note="hypothetical protein"
    /codon_start=1
    /transl_table=11

product="Vng2148h"
/protein_id="AAG20285.1"
/db_xref="GI:10581564"
/translation="MSRPLAVALVAVVATAMATGLAAGPPTASQSPAVGSPDT
TVRLTSADAAAFGTSPALSVGTGLAAGATGISTAYLERLDTLASVANTDAERARI
RTELDWATNRTYVQERDARDAYAAAGTTTTEFLTVGVSVHAAATRIEGLGPPD
NRALYTLADSYNLNLTTRISOVNRLQTVQGRERLAADVGDRAVHVTVGDMV
LFGDGTIVRETVPGNLDAATADYGDPIAFIGATVPWKHNRSGVSYSLIGQYAG
LFSHSDHGEISLYDSTHVFVERQTRLPOTPSYTTVNAATDTLTLSVEYVAGG
FVNKRVETASGTPVDITISVNGIVGTIGSDGRWVFSVGPAGEYVAAIVGDRVTTV
TADPHPPAGPNTTTLPR"
2566..3729
    /gene="VNG2149H"
    /note="hypothetical protein"
    /codon_start=1
    /transl_table=11
    /product="Vng2149h"
    /db_xref="GI:10581565"
    /translation="MRQALLLVALACLVAVPVGGAGGHPAPTAETTAMPAATGSG
PQSYVDITVETDGDHAWNYSATYPLGDSGVAFORVSTAFAGETSAGFSAETFK
AAVPGASRVGREMSVIDVRRASRIEVDHNETGVLLIRFTWTFPAADVNTIYDAF
SGSFGDLRAGQQLRVPPENYQADVSPNTNVGGAYQWEGGQAFADGSLVFPVS
PGLSGTGTIPATVGAADVLLGAAGTAWAYSRTLTLPFGTGGDSEGAFTAVTGDG
GESAPTADDPDDGDDSAEREDRIDPELLSDEERVELITDNGRMRKQSRIVE
ETRWSTAKVSQVLSGDMDEGRVKEKLIRIGRENLI SLPGEGVSDTLAGEHDETE"
3799..4599
    /gene="etfb"
    /note="synonym: VNG2150G"
3799..4599
    /gene="etfb"
    /note="etfb"
    /codon_start=1
    /transl_table=11
    /product="electron transfer flavoprotein subunit beta"
    /protein_id="AAG20288.1"
    /db_xref="GI:10581566"
    /translation="MKVLVTVKEVAIVEDDFEISGTEISTEYLDYDLNEDDYAVEEA
VQLADAGDQVEVAVTIGPERAETIRMALAKADRAVVRVWDDAIETELLDVETKAR
LLAAVDEDEDDELVPVSGVNDGSGATGIALAETLGVQAAVNDLDTAGVLDEGVA
SVRLEGGVELDDELPAVLTQTLGNEPRYASLRGIRAOQSKEIAPKSLDLDGLD
ADIVASSDITANYEPETESDAEYIEGDAGEQASGVADVLDAGVWGE"
4596..5564
    /gene="etfa"
    /note="synonym: VNG2151G"
4596..5564
    /gene="etfa"
    /note="etfa"
    /codon_start=1
    /transl_table=11
    /product="electron transfer flavoprotein subunit alpha"
    /protein_id="AAG20288.1"
    /db_xref="GI:10581567"
    /translation="MSVLAVADHRRGELRGVSEFLLTAGRELAMDTELHAAVVSVD
TDGFADELNRAGVDHVAHEGAEPNDVYQVAAVAGLDLDAVVLTPNSVNGLDYA
PAVATRTYVSDAIGDHTLEVTREMGSKVETTWESDGYVYVSGIRGAEPAA
EGVGDAEISTFDADIDEAIGATVNGFEVGGVDVADIADAEFLVSGVSGIEEENIAL
IEALAEITGATLSSSRPIVDNGLAKNRQVGSQKQVTPDYVLVIGISGAVQHVAGMK
GADATIAVNTDPNAPIFDIADYIVGDLFVVPVPALEIEFGGEAPSV"
    complement (5588..6238)
    /gene="VNG2152C"
    complement (5588..6238)
    /gene="VNG2152C"
    /note="conserved hypothetical protein"
    /codon_start=1
    /transl_table=11
    /product="Vng2152c"
    /protein_id="AAG20289.1"
    /db_xref="GI:10581568"
    /translation="MRGQPVHRRSVLALVGGGAVSALAGCTDITDGDATTTTPTPR
SETQPTATDDIPIAGSPDFGGLGASNAFADGAALPDRYTRSGADVNPPLSTAGV"

```



```
/gene="OJ1065_B06.1"
/notes="supported by full-length cDNA(s): AK105640"
1766..3334
/gene="OJ1065_B06.1"
/notes="contains full-length cDNA(s): AK105640
similar to Arabidopsis thaliana chromosome 2, NP_850414"
/codon_start=1
/product="unknown protein"
/protein_id="BAC83058.1"
/db_xref="GI:34393497"
/translation="MLTGLSFRVSVDGFGAKGSDLGLEVRQVTTFAKWTSGDGG
CATPRPQVAGSALGHYAQRNDRLPGHVLVAROVSGGGLLAPLPLDVPATSSOD
DAHFWLTPDGGRAIGAVATASPDKPREDEACVADFDADACEATVWDKDGPSA
LALPAVGVARGVHAGTFLVARSADATASALACLNKNGAAYTSCMPDLAVNAL
LAAPVQFLHEDPELFSVTFWFGALLYQKSGTPTFVAADGNSLPGCGNDGG
YLDLPVDFORERVKGDLAGAKYVQAKEMLGATADLAVWFYFNFNGPARAKVGP
LTIPLKGTGDEWHVTLRVSNFSGBELLMYFSOSAGWADAPOLEYLDGNGRPS
AVSSIHGALVPRAGLVQGDARLGVIRNDCDRGLDTGAGRCVSAEYLGSGS
GGVAPPTWLLDRENGPREEDYIGREINVAKLIPRSTRERLKLVSFVGEQPTGP
RMKGSWRNDRERAK"
4131..11355
/gene="OJ1065_B06.2"
/translation="4131..4486,5711..5834,5914..6058,6190..6415,
7021..7094,7268..7360,8691..8808,9598..9660,10032..10064,
10583..10659,10807..11355"
/gene="OJ1065_B06.2"
/notes="supported by full-length cDNA(s): AK072012"
Join(4248..4486,5711..5834,5914..6058,6190..6415,
7021..7094,7268..7360,8691..8808,9598..9660,10032..10064,
10583..10659,10807..11354)
/gene="OJ1065_B06.2"
/notes="supported by full-length cDNA(s): AK102382"
Join(4298..4486,5711..5834,5914..6058,6190..6415,
7021..7094,7268..7360,8691..8808,9598..9660,10032..10064,
10583..10659)
/gene="OJ1065_B06.2"
/notes="contains full-length cDNA(s):
AK072012, AK102382, AK063129"
/codon_start=1
/product="putative tocopherol polyphenyltransferase"
/protein_id="BAC83059.1"
/db_xref="GI:34393498"
/translation="MASLASPLPCRAAATASRGRPAFRLGPPPPASPILLSASA
RPPALNAARSDRVRVCSQAGAPAPLSTLSDKSCWRFLRPIRTGALG
SMILVARALPOLINLWLVKAFYGLVALICNGYIVGINOYDIRIDKNKPYLP
TAAGDLVQTAWLAVLPAAGFSIVTNFGPFTSLXGLGFGLTIVSVPEFLKRY
PVAELIATVRGFLNFGVYATRAALGLTFQSSPVAFITCFVTLFALVIATKDL
PDVGGRIQSLATKQGVNRNIAPLGSGLLIANYAAIAVAFMLPQAFRTVWVPVH
RALAVGIIFQWVLEQAKYTKDAISQYRFIWNLFYAEYIFFLI"
complement(join(13924..14038,14095..14147,14201..14275))
/gene="OJ1065_B06.3"
complement(join(13924..14038,14095..14147,14201..14275))
/gene="OJ1065_B06.3"
/notes="hypothetical ORF
predicted by GeneMark.hmm
this category is not included in IRGSP standard"
16161..20833
/gene="OJ1065_B06.4"
Join(<16161..16612,19790..20147,20207..>20833)
/gene="OJ1065_B06.4"
/notes="start and end point are not identified"
Join(16161..16612,19790..20147,20207..20833)
/gene="OJ1065_B06.4"
/codon_start=1
/product="putative Nt-gh3 deduced protein"
/protein_id="BAC83060.1"
/db_xref="GI:34393499"
/translation="MAAGKSTIGASRLSSLSMAAKEPDVFNRLRIELTNSVDA
VQVRLAIIIGNDAEYLDKGLDASDTDRATFRKVPVASYDDLKPKVRIANGOR
SPILSTHPIIEFTSSGAGERKLMPITVDEMARREVLSLATSVLNVYVGLHTGK
GLYFLFARSEKTPGGLTAQALTSYKSEHFKAAYVTSMAAILCEDASQSMYAQ
LCGLCQRHDVLRGVAFVAAALVRAIRFLQNLWAOLADTETGLNPRVTSDECSR
GDWTGIVRELKPKCLNVVTVGWMAQIPIPLQYVSGGLPIVSGWVASECFEGLNLR
```

```
PLCGPSEVYTIMPTAYFEFLPVGEAVDASNLVELARVEDGREVEVVVTVVAGLNRY
RVDVLCVTGFHNAAPQFRFVRQSVLLSIEADKDEAELQRAVERASSALLRPRGAS
IVEYTSRACRTERVGFHSPALPHWTLP"
complement(join(21777..21922,22075..22156))
/gene="OJ1065_B06.5"
complement(join(21777..21922,22075..22156))
/gene="OJ1065_B06.5"
/notes="hypothetical ORF
predicted by GlimmerM
this category is not included in IRGSP standard"
22591..23165
/gene="OJ1065_B06.6"
Join(<22591..22692,23007..>23165)
/gene="OJ1065_B06.6"
/notes="start and end point are not identified"
Join(22591..22692,23007..23165)
/gene="OJ1065_B06.6"
/notes="predicted by GENSCAN etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAC83061.1"
/db_xref="GI:34393500"
/translation="MSGHGCADCWVMPCRIGTLCOGRGPAQAARNSRQTSAMHALL
ASRVAPHTTSPPPQLTPSPVSPQSGERTSSGRSTMAGPG"
complement(23923..24723)
/gene="OJ1065_B06.7"
complement(join(<23923..24111,24601..>24723))
/gene="OJ1065_B06.7"
/notes="start and end point are not identified"
complement(join(23923..24111,24601..24723))
/gene="OJ1065_B06.7"
/notes="predicted by FGENSEH etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAC83062.1"
/db_xref="GI:34393501"
/translation="MGAQILFVLSLFPPLSPSLASNDGSGPGMARVYRDGDTATT
RQVDICILHGTTTKRAAGKKAAMQECASPPANYFKNKADYNAASASAGAINY
K"
25126..26796
/gene="OJ1065_B06.8"
Join(<25126..25183,25708..26045,26134..>26796)
/gene="OJ1065_B06.8"
/notes="start and end point are not identified"
Join(25126..25183,25708..26045,26134..26796)
/gene="OJ1065_B06.8"
/notes="predicted by GeneMark.hmm etc."
/codon_start=1
/product="hypothetical protein"
/protein_id="BAC83063.1"
/db_xref="GI:34393502"
/translation="MAQPNDOHNDCTPTSPSGIROWAGQAGGSGSAAGDGRGD
GGSQHDIPSPDAPKQSTPPTTSPASRLQCNTHIGAKTSNAINGVFVLLLFQ
VTFVSLGGGGLVIVGGDGAAREDEPSLPDPSLRQKWDVLLAYAQIALAVDRA
TDVGIKTACGSFQSAAGAFAMWRESGVDTAAVAGATTIDVLSLLSPSPNHLVLP
```

Query Match 89.5%; Score 17; DB 8; Length 113170;  
Best Local Similarity 100.0%; Pred. No. 5.1e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TCGACGCCGCCGATGAC 18  
|||||  
Db 72265 TCGACGCCGCCGATGAC 72249

RESULT 22  
AP003013/c  
LOCUS AP003013 347950 bp DNA linear BCT 15-MAY-2001  
DEFINITION Mesorhizobium loti DNA, complete genome, section 20/21.  
ACCESSION AP003013 BA000012  
VERSION AP003013.2 GI:14027324  
KEYWORDS Mesorhizobium loti  
SOURCE







```

/codon_start=1
/transl_table=11
/protein_id="BAB53602.1"
/db_xref="GI:14027333"
/translation="MNQAKRLRHSHSACPHDCPSTCALEVELLDGNRIKRVHGAKANT
YTGVCVAKVAVRVDHEDRLKPLIRAGAKGDGVQWQASWEALDGLVAERFIAAE
YKSGEYVFPYFAASCIATVAGSVOYEGGAFHNSGIFKLNQVLEGRMRDPSIR
YLDHSRIGPVLITGASDALYGGPPVTAMLIONTPNVNAPCEKLVKOGFLRDDLLFACVH
EQMTDARLADVLPAATMEHDDVYKGGNOHITLGPKLIEPPSGPRTHHFIQSL
AERGVADRFGGLTEQOHDIIILKRGGLGSPSLKEQKQVLDLQPDFFTAHFYKFGSH
SDGFRPADWTQANRPPKSMGLFPGVPERLPPDHDVLEAVGQVDRVFNKFLITSP
ARNFLNFTFASTPVSKERGPALLDHPDADLGLADGRVEVGNQGRDLVHLKFF
DGIKGVVIAEGVWPNSEHGERGINVLTGADAPAPYGGAAFDHKNVWLRKAG"
complement(8203..9390)
/gene="ml17934"
complement(8203..9390)
/notes="similar to dimethylalanine monooxygenase"
/codon_start=1
/transl_table=11
/protein_id="BAB53603.1"
/db_xref="GI:14027334"
/translation="MFLTSGRELSREARLDDTTTVAIIGAGPAGLAVACLQAGVD
FIIEKQQAAPWARRHYERVHLHTTKRYSLSLFPVPPKRYIPRPRALFVLDIAYA
QRDLRFQGTQKAVTODGKRWVDAAGFLRAKDVVIASGYNABELPFAFADITF
TGKTLHGADYNNARLPFAQSIVLTGMNTGAEIALDLAENGAQPTISVRGGVHIVPRE
LFGVPIQWVGNAALGFLQRLNDALFPVILDLVLGRDKYGLKRPFGQCLLEQIIVASRI
PVIDVGTGKIREGAIKVAPDITEISORGARFADGKHGFDAIIFATGVRPGVAFLE
PGIQDRSGVTPKASDILGLYLIGHNAVTLGLREIGIEAQAIAADDIHRHLNRKKAEL
LPV"
complement(9426..9905)
/gene="ml17935"
complement(9426..9905)
/notes="similar to dimethylalanine monooxygenase"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53604.1"
/db_xref="GI:14027335"
/translation="MPPSPSPPLDSFDLAILVLQDNTTPQRLIGEAVNLSPAPVQR
IKRQAGVIAASNVAIIEPAAGKPTIFVEVELESERTELIDAAKQFQSTQTEVQOC
YVYTGADFLIVITVADMGAYEALTRKLFEGSNVRKFRFVAVDRVKVGLTVPFLPD"
10133..11218
/gene="ml17936"
10133..11218
/gene="ml17936"
/notes="hypothetical protein"
/codon_start=1

Query Match 89.5%; Score 17; DB 1; Length 347950;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGA 17
|||||
Db 198925 GTCGACGCCGCCGATGA 198909

RESULT 23
AY023290/c
LOCUS 227 bp DNA linear PLN 07-FEB-2001
DEFINITION Oryza sativa microsatellite MRS5615 containing (GCG)X9, closest to
marker G148, genomic sequence.
ACCESSION AY023290
VERSION AY023290.1 GI:12706506
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriactoidae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 227)
AUTHORS Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.

/codon_start=1
/transl_table=11
/protein_id="BAB53602.1"
/db_xref="GI:14027333"
/translation="MNQAKRLRHSHSACPHDCPSTCALEVELLDGNRIKRVHGAKANT
YTGVCVAKVAVRVDHEDRLKPLIRAGAKGDGVQWQASWEALDGLVAERFIAAE
YKSGEYVFPYFAASCIATVAGSVOYEGGAFHNSGIFKLNQVLEGRMRDPSIR
YLDHSRIGPVLITGASDALYGGPPVTAMLIONTPNVNAPCEKLVKOGFLRDDLLFACVH
EQMTDARLADVLPAATMEHDDVYKGGNOHITLGPKLIEPPSGPRTHHFIQSL
AERGVADRFGGLTEQOHDIIILKRGGLGSPSLKEQKQVLDLQPDFFTAHFYKFGSH
SDGFRPADWTQANRPPKSMGLFPGVPERLPPDHDVLEAVGQVDRVFNKFLITSP
ARNFLNFTFASTPVSKERGPALLDHPDADLGLADGRVEVGNQGRDLVHLKFF
DGIKGVVIAEGVWPNSEHGERGINVLTGADAPAPYGGAAFDHKNVWLRKAG"
complement(8203..9390)
/gene="ml17934"
complement(8203..9390)
/notes="similar to dimethylalanine monooxygenase"
/codon_start=1
/transl_table=11
/protein_id="BAB53603.1"
/db_xref="GI:14027334"
/translation="MFLTSGRELSREARLDDTTTVAIIGAGPAGLAVACLQAGVD
FIIEKQQAAPWARRHYERVHLHTTKRYSLSLFPVPPKRYIPRPRALFVLDIAYA
QRDLRFQGTQKAVTODGKRWVDAAGFLRAKDVVIASGYNABELPFAFADITF
TGKTLHGADYNNARLPFAQSIVLTGMNTGAEIALDLAENGAQPTISVRGGVHIVPRE
LFGVPIQWVGNAALGFLQRLNDALFPVILDLVLGRDKYGLKRPFGQCLLEQIIVASRI
PVIDVGTGKIREGAIKVAPDITEISORGARFADGKHGFDAIIFATGVRPGVAFLE
PGIQDRSGVTPKASDILGLYLIGHNAVTLGLREIGIEAQAIAADDIHRHLNRKKAEL
LPV"
complement(9426..9905)
/gene="ml17935"
complement(9426..9905)
/notes="similar to dimethylalanine monooxygenase"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53604.1"
/db_xref="GI:14027335"
/translation="MPPSPSPPLDSFDLAILVLQDNTTPQRLIGEAVNLSPAPVQR
IKRQAGVIAASNVAIIEPAAGKPTIFVEVELESERTELIDAAKQFQSTQTEVQOC
YVYTGADFLIVITVADMGAYEALTRKLFEGSNVRKFRFVAVDRVKVGLTVPFLPD"
10133..11218
/gene="ml17936"
10133..11218
/gene="ml17936"
/notes="hypothetical protein"
/codon_start=1

Query Match 89.5%; Score 17; DB 1; Length 347950;
Best Local Similarity 100.0%; Pred. No. 4.6e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGA 17
|||||
Db 198925 GTCGACGCCGCCGATGA 198909

RESULT 23
AY023290/c
LOCUS 227 bp DNA linear PLN 07-FEB-2001
DEFINITION Oryza sativa microsatellite MRS5615 containing (GCG)X9, closest to
marker G148, genomic sequence.
ACCESSION AY023290
VERSION AY023290.1 GI:12706506
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriactoidae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 227)
AUTHORS Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.

```

```

TITLE Simple sequence repeats from Monsanto rice genomic sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 227)
AUTHORS Tao,N., Barbazuk,W.B., Liu,J., Wu,K. and Barry,G.F.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2001) Genomics, Monsanto, 800 North Lindbergh
Blvd., Creve Coeur, MO 63167, USA
COMMENT Derived from rice genomic sequences generated from the Monsanto
Rice Genome Sequencing project. Please see
http://www.rice-research.org for more information. The sequence
data were produced primarily in the laboratories of Dr. Leroy Hood
at the University of Washington in Seattle.
FEATURES
source
1..227
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4330"
repeat_region
1..227
/notes="microsatellite MRS5615"
/rpt_type=tandem
/rpt_unit="ggc"
ORIGIN
Query Match 86.3%; Score 16.4; DB 8; Length 227;
Best Local Similarity 94.4%; Pred. No. 1.4e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGACGCCGCCGATGAC 18
|||||
Db 41 GTCGACGCCGCCGAGGAC 24

RESULT 24
AFAAZU/c
LOCUS 810 bp DNA linear BCT 24-APR-1993
DEFINITION A. denitrificans azurin (azu) gene, complete cds.
ACCESSION X30388
VERSION X30388.1 GI:141901
KEYWORDS azurin.
SOURCE Achromobacter xylosoxidans subsp. denitrificans
ORGANISM Achromobacter xylosoxidans subsp. denitrificans
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Alcaligenaceae; Achromobacter.
REFERENCE 1 (bases 1 to 810)
AUTHORS Hoitink,C.W., Woudt,L.P., Turenhout,J.C., van de Kamp,M. and
Canter,G.W.
TITLE Isolation and sequencing of the Alcaligenes denitrificans
azurin-encoding gene: comparison with the genes encoding blue
copper proteins from Pseudomonas aeruginosa and Alcaligenes
faecalis
JOURNAL Gene 90 (1), 15-20 (1990)
MEDLINE 90337337
PUBMED 2116366
COMMENT Original source text: A.denitrificans (strain NCTC9582) DNA.
Draft entry and computer-readable sequence for (1) kindly submitted
by G.W.Canter, 01-DEC-1989, for release after publication.
FEATURES
source
1..810
/organism="Achromobacter xylosoxidans subsp.
denitrificans"
/mol_type="genomic DNA"
/sub_species="denitrificans"
/db_xref="taxon:32002"
190..203
/notes="fmr-box"
251..267
/notes="ntrA-box"
296..299
/notes="ribosome binding site"
307..756
/notes="azurin (azu) precursor"
/codon_start=1
/transl_table=11
misc_feature
251..267
/notes="ntrA-box"
misc_feature
296..299
/notes="ribosome binding site"
RBS
307..756
/notes="azurin (azu) precursor"
CDS
/codon_start=1
/transl_table=11

```

```

/protein_id="AA21954.1"
/db_xref="GI:141902"
/translation="MLAKATLAIVLSAASLPVLAQCEATIESNDAMQYNLKENWVDK
SKQFTVHLKHVGVKAVGHNVLTREADKQGVATDGMNAGLAQDYVKAQDTRVIA
HTKVIIGGESDSTVTFDVSGLTPEAYAYFCSPGHAWMKGTLSLN"
sig_peptide
307..366
/note="azurin signal peptide"
mat_peptide
367..753
/product="azurin"
misc_signal
778..800
/note="terminator"

ORIGIN
Query Match 86.3%; Score 16.4; DB 1; Length 810;
Best Local Similarity 94.4%; Pred. No. 1.3e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCGACGCCGCCGATGACC 19
Db 638 TCGCGCGCGCGATGACC 621

RESULT 25
AX653785/c
LOCUS AX653785 963 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 3655 from Patent WO03000898.
ACCESSION AX653785
VERSION AX653785.1 GI:29156599
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Katagiri,F., Quan,S., Tao,X., Whitcham,S., Xie,Z., Zhu,T. and Zou,G.
Plant genes involved in defense against pathogens
Patent: WO 03000898-A 3655 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
Source
1..963
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 86.3%; Score 16.4; DB 6; Length 963;
Best Local Similarity 94.4%; Pred. No. 1.3e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGAC 18
Db 501 GTCGCGCGCGCGATGAC 484

RESULT 26
AX660080/c
LOCUS AX660080 1077 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 437 from Patent WO03000906.
ACCESSION AX660080
VERSION AX660080.1 GI:29162039
KEYWORDS
SOURCE
ORGANISM
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 Glazebrook,J., Briggs,S., Cooper,B., Goff,S.A., Moughamer,T.,
Katagiri,F., Kreis,J., Provart,N., Rieke,D. and Zhu,T.
Plant disease resistance genes
Patent: WO 03000906-A 437 03-JAN-2003;

```

```

FEATURES
Source
1..1077
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 86.3%; Score 16.4; DB 6; Length 1077;
Best Local Similarity 94.4%; Pred. No. 1.3e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGACGCCGCCGATGAC 18
Db 171 GTCGACGTCGCGATGAC 154

RESULT 27
AB025282/c
LOCUS AB025282 1339 bp mRNA linear VET 14-APR-1999
DEFINITION Gallus gallus cpgc mRNA for pepsinogen C, complete cds.
ACCESSION AB025282
VERSION AB025282.1 GI:4589839
KEYWORDS pepsinogen C.
SOURCE Gallus gallus
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (sites)
Sakamoto,N., Saiga,H. and Yasugi,S.
Analysis of temporal expression pattern and cis-regulatory
sequences of chicken pepsinogen A and C
Biochem. Biophys. Res. Commun. 250 (2), 420-424 (1998)
JOURNAL
MEDLINE 98440813
PubMed 9753645
REFERENCE
2 (bases 1 to 1339)
Sakamoto,N. and Yasugi,S.
Direct Submission
Submitted (24-MAR-1999) Sadao Yasugi, Tokyo Metropolitan
University, Department of Biology, The Graduate School of Science;
Minamiohsawa 1-1, Hachioji, Tokyo 192-0397, Japan
(E-mail: yasugi-sadao@ec.metro-u.ac.jp, Tel: +81-426-77-2572,
Fax: +81-426-77-2559)
FEATURES
Source
1..1339
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
1..1339
/genes="cpgc"
20..1189
/genes="cpgc"
/codon_start=1
/product="pepsinogen C"
/protein_id="BAA76893.1"
/db_xref="GI:4589840"
/translation="MKRLTLVLCILHLCGLRVLKKGKSTREAKESGVLHDYLAN
HRYDPAYKFNFNATAYEPLANNMDSYGIISIGTPQNFVLPTDGSNLAWPST
LQSQCAHNHFDNPSSTFTQDEFFSLQSGSLTGICFGFDVTICGSIITNQEF
GLSETEFGTSFLYSPFDGLAFPSISAGGATVQMKVLQNLJDFPVFSYLSGQE
GSGQGLVFGVDPNLYQTITWPTQTYWQIGIEDFAVGQSSGWCSCQCGQIVD
TGTSLTVPNQVFTLMQYIGAADSGQYVASCNIEYMTITFVIGTSPLPPSA
YMLQNSDICTVGIESTYLPSTGQPLWILGDVFLRVYSYIDMGNQVGFATAV"
1339
/genes="cpgc"
/note="9 A nucleotides"

polyA_site
1339

ORIGIN
Query Match 86.3%; Score 16.4; DB 5; Length 1339;
Best Local Similarity 94.4%; Pred. No. 1.2e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 GTCGACGCCGCCGATGAC 18
Db 712 GTCGACGCCGCCGAGAC 595

AK104317 1389 bp mRNA linear PLN 24-JUL-2003
Oriza sativa (japonica cultivar-group) cDNA clone:001-024-C03, full
insert sequence.
ACCESSION AK104317
VERSION AK104317.1 GI:32989526
FLI CDNA: oligo capping.
KEYWORDS Oriza sativa (japonica cultivar-group)
SOURCE Oriza sativa (japonica cultivar-group)
ORGANISM Oriza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 The Rice Full-length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-length cDNA Project Team,
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Hashizume, W., Yamada, H., Hotta, I.,
Hirakawa, K., Yamashita, Y., Hayatsu, N., Hiramoto, K., Hirakawa, T.,
Horii, F., Hotta, I., Iida, J., Ikehara, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M.,
Kagawa, S., Kato, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Koda, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokemizo, S. and
Yoshimura, A.
Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agricultural Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan [E-mail: skkuch@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007]
This clone is one of the 28k full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,

RESULT 29
AK104317/c
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) cDNA clone:001-024-C03, full
insert sequence.
ACCESSION AK071251.1 GI:32981274
VERSION AK071251
KEYWORDS CAP trapper.
SOURCE Oriza sativa (japonica cultivar-group)
ORIGIN
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 The Rice Full-length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-length cDNA Project Team,
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Hashizume, W., Yamada, H., Hotta, I.,
Hirakawa, K., Yamashita, Y., Hayatsu, N., Hiramoto, K., Hirakawa, T.,
Horii, F., Hotta, I., Iida, J., Ikehara, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M.,
Kagawa, S., Kato, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Koda, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokemizo, S. and
Yoshimura, A.
Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agricultural Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan [E-mail: skkuch@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007]
This clone is one of the 28k full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,

FEATURES
source
1..1389
/organism="Oriza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-024-C03"

ORIGIN
Query Match 86.3%; Score 16.4; DB 8; Length 1389;
Best Local Similarity 94.4%; Pred. No. 1.2e+04;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GTCGACGCCGCCGATGAC 18
Db 608 GTCGACGCCGCCGATGAC 591

RESULT 29
AK071251/c
LOCUS
DEFINITION
Oriza sativa (japonica cultivar-group) cDNA clone:001-024-C03, full
insert sequence.
ACCESSION AK071251.1 GI:32981274
VERSION AK071251
KEYWORDS CAP trapper.
SOURCE Oriza sativa (japonica cultivar-group)
ORIGIN
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 The Rice Full-length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-length cDNA Project Team,
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Hashizume, W., Yamada, H., Hotta, I.,
Hirakawa, K., Yamashita, Y., Hayatsu, N., Hiramoto, K., Hirakawa, T.,
Horii, F., Hotta, I., Iida, J., Ikehara, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M.,
Kagawa, S., Kato, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Koda, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokemizo, S. and
Yoshimura, A.
Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agricultural Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan [E-mail: skkuch@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007]
This clone is one of the 28k full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,

```

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice  
Science 301 (5631), 376-379 (2003)  
22752273  
12869764  
2 (bases 1 to 1502)  
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K., Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K., Imoto,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M., Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M., Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Kurotaki,T., Kusumegi,T., Li,C., Lu,M., Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A., Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M., Namiki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K., Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H., Osato,N., Ota,Y., Otsu,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K., Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S., Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W., Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and Yoshimura,A.  
Direct Submission  
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kanondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skkikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)  
This clone is one of the 28K full-length cDNA clones from japonica rice.  
URL : http://cdna01.dna.affrc.go.jp/cDNA/  
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T., and Yamamoto,M.  
FAIS Genome Sequencing & Analysis Group: Otsu,Y., Iida,Y., Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M., Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Narikawa,R., Nishikura,J., Oka,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.  
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayatsu,N., Hiramoto,K., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.  
Location/Qualifiers  
1..1502  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="J023087H21"

TITLE  
JOURNAL

Query Match 86.3%; Score 16.4; DB 8; Length 1502;  
Best Local Similarity 94.4%; Pred. No. 1.2e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

FEATURES  
source  
1..1502  
/organism="Oryza sativa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4530"

ORIGIN  
Query Match 86.3%; Score 16.4; DB 6; Length 1560;  
Best Local Similarity 94.4%; Pred. No. 1.2e+04;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGACGCGCGCGCGATGAC 18  
|||||  
DB 624 GTGCGCGCGCGCGATGAC 607

RESULT 30  
AX366999/c  
LOCUS AX366999 1560 bp DNA linear PAT 16-FEB-2002  
DEFINITION Sequence 26 from Patent WO0198509.  
ACCESSION AX366999  
VERSION AX366999.1 GI:18698275  
KEYWORDS  
SOURCE Oryza sativa  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
1  
REFERENCE  
AUTHORS Lanahan,M.B., Desai,N.M. and Gasdaska,P.Y.  
TITLE Grain processing method and transgenic plants useful therein  
JOURNAL Patent: WO 0198509-A 26 27-DEC-2001;  
SYNOPSIS Syngenta Participations AG (CH)  
FEATURES  
source 1..1560  
/organism="Oryza sativa"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:4530"

QY 1 GTGACGCGCGCGCGATGAC 18  
|||||  
DB 906 GTGCGCGCGCGCGATGAC 889

Search completed: June 20, 2004, 11:43:55  
Job time : 531.915 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 678.836 Seconds  
(without alignments)  
1596.226 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccggagggccagagggcacatagc 25

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg.hum.\*

31: em.htg.inv.\*

32: em.htg.other.\*

33: em.htg.mus.\*

34: em.htg.pln.\*

35: em.htg.rhod.\*

36: em.htg.man.\*

37: em.htg.vrt.\*

38: em.sy.\*

39: em.htgo.hum.\*

40: em.htgo.mus.\*

41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	15619	1	AE006916	AE006916 Mycobacte
2	25	100.0	341957	15	BX842572	BX842572 Mycobacte
3	25	100.0	343050	1	BX248334	BX248334 Mycobacte
4	24	96.0	38721	1	MSGY219	AD000013 Mycobacte
5	20.2	80.8	60740	2	AC016557	AC016557 Homo sapi
6	20.2	80.8	168008	9	AC008544	AC008544 Homo sapi
7	20.2	80.8	171004	2	AC034202	AC034202 Homo sapi
8	20.2	80.8	183734	9	AC109479	AC109479 Homo sapi
9	20.2	80.8	185329	2	AC023587	AC023587 Homo sapi
10	20.2	80.8	199768	2	AC136594	AC136594 Homo sapi
11	20.2	80.8	220632	2	AC136596	AC136596 Homo sapi
12	19.2	76.8	104526	9	AC114743	AC114743 Homo sapi
13	19.2	76.8	146868	2	AC016802	AC016802 Homo sapi
14	19.2	76.8	166558	2	AC112141	AC112141 Homo sapi
15	19.2	76.8	169651	2	AC084771	AC084771 Homo sapi
16	18.6	75.2	258723	2	AC110113	AC110113 Rattus no
17	18.6	74.4	1276	10	WU93191	U93191 Mus musculu
18	18.6	74.4	59817	6	AX695375	AX695375 Sequence
19	18.6	74.4	60042	9	HSIMMDL	AC100484 Mus muscu
20	18.6	74.4	92588	9	HSIMMDL	X97051 DNA sequenc
21	18.6	74.4	100057	10	AL929178	AL929178 Mouse DNA
22	18.6	74.4	145052	10	AL603889	AL603889 Mouse DNA
23	18.6	74.4	149321	2	AL731845	AL731845 Mus muscu
24	18.6	74.4	151555	10	AL732474	AL732474 Mouse DNA
25	18.6	74.4	157090	9	AB019441	AB019441 Homo sapi
26	18.6	74.4	157361	10	AL645923	AL645923 Mouse DNA
27	18.6	74.4	165116	4	AC092729	AC092729 Canis fam
28	18.6	74.4	165588	9	AL590966	AL590966 Human DNA
29	18.6	74.4	169802	9	CNS01DT2	AL122127 Human chr
30	18.6	74.4	183564	10	BX546505	BX546505 Mouse DNA
31	18.6	74.4	183786	10	AL607087	AL607087 Mouse DNA
32	18.6	74.4	190512	2	AC140278	AC140278 Mus muscu
33	18.6	74.4	190981	2	BX890380	BX890380 Mus muscu
34	18.6	74.4	191941	2	AC102693	AC102693 Mus muscu
35	18.6	74.4	191969	2	AC101974	AC101974 Mus muscu
36	18.6	74.4	195848	2	AC102615	AC102615 Mus muscu
37	18.6	74.4	200088	10	AC078897	AC078897 Mus muscu
38	18.6	74.4	202848	2	AC113503	AC113503 Mus muscu
39	18.6	74.4	203592	10	AL669921	AL669921 Mouse DNA
40	18.6	74.4	210369	2	AC109502	AC109502 Mus muscu
41	18.6	74.4	210919	2	AC115554	AC115554 Rattus no
42	18.6	74.4	232278	2	AC087876	AC087876 Mus muscu
43	18.6	74.4	237282	10	AL772371	AL772371 Mouse DNA
44	18.6	74.4	239525	2	AC095150	AC095150 Rattus no
45	18.6	74.4	241082	2	AC127158	AC127158 Rattus no
46	18.6	74.4	242224	2	AC105066	AC105066 Mus muscu
47	18.6	74.4	243659	2	AC128597	AC128597 Rattus no
48	18.6	74.4	244903	2	AC097807	AC097807 Rattus no
49	18.6	74.4	257471	2	AC139675	AC139675 Mus muscu
50	18.6	74.4	267943	2	AC136702	AC136702 Mus muscu
51	18.6	74.4	309350	1	AP005215	AP005215 Corynebact
52	18.6	74.4	309804	1	AE016865	AE016865 Pseudomon
53	18.6	73.6	1095	9	HMMHRC01	M23903 Human MHC c
54	18.4	73.6	14646	9	HSIHLASBA	X03100 Human HLA-S
55	18.4	73.6	14782	9	HSIHLADFB	X02228 Human HLA-D
56	18.4	73.6	64380	9	AL805913	AL805913 Human DNA
57	18.4	73.6	106728	9	AL805913	AL805913 Human DNA
58	18.4	73.6	124899	9	AL645931	AL645931 Human DNA
59	18.4	73.6	181228	2	AC011086	AC011086 Homo sapi
60	18.4	73.6	187964	9	AL662824	AL662824 Human DNA
61	18.4	73.6	196572	2	AC118785	AC118785 Rattus no
62	18.4	73.6	243279	2	AC099068	AC099068 Rattus no
63	18.4	73.6	272405	2	AC094363	AC094363 Rattus no
64	18.2	72.8	1397	3	AY069277	AY069277 Drosophil
65	18.2	72.8	2140	3	DMCOLT	Y12495 D.melanogas

66	18.2	72.8	32327	10	AL627068	AL627068 Mouse DNA	C 139	17.6	71.2	267398	2	AC106519	AC106519 Rattus no
67	18.2	72.8	35734	2	AC019974	AC019974 Drosophil	140	17.6	70.4	637	6	AX229939	AX229939 Sequence
68	18.2	72.8	68398	2	AC131045	AC131045 Homo sapi	C 141	17.6	70.4	1152	6	AX647147	AX647147 Sequence
69	18.2	72.8	90442	6	AX695968	AX695968 Homo sapi	C 142	17.6	70.4	1351	10	BC038881	BC038881 Mus muscu
70	18.2	72.8	124700	3	AC005558	AC005558 Drosophil	C 143	17.6	70.4	3103	10	BC062653	BC062653 Mus muscu
71	18.2	72.8	150720	9	AL513016	AL513016 Human DNA	C 144	17.6	70.4	6081	10	BC043114	BC043114 Mus muscu
72	18.2	72.8	179125	3	AC009392	AC009392 Drosophil	145	17.6	70.4	6450	3	CEU36199	CEU36199 Caenorhabdi
73	18.2	72.8	190372	2	AL928969	AL928969 Mus muscu	146	17.6	70.4	21733	3	CEM10055	CEM10055 Caenorhabdi
74	18.2	72.8	190485	10	AL713913	AL713913 Mouse DNA	C 147	17.6	70.4	35210	5	AY170464	AY170464 Rattus no
75	18.2	72.8	204926	2	AC128263	AC128263 Rattus no	C 148	17.6	70.4	36384	9	AC004178	AC004178 Homo sapi
76	18.2	72.8	216069	2	AC122086	AC122086 Rattus no	C 149	17.6	70.4	38476	9	U73024	U73024 Homo sapien
77	18.2	72.8	228910	2	AC109420	AC109420 Rattus no	C 150	17.6	70.4	53675	9	AL391708	AL391708 Human DNA
78	18.2	72.8	231989	10	AL805897	AL805897 Mouse DNA	C 151	17.6	70.4	55748	2	AC105912	AC105912 Homo sapi
79	18.2	72.8	244389	2	AC111858	AC111858 Rattus no	C 152	17.6	70.4	61843	8	NC15E6	NC15E6 Neurospor
80	18.2	72.8	247936	2	AC112533	AC112533 Rattus no	C 153	17.6	70.4	72305	8	NCB1385	NCB1385 Neurospor
81	18.2	72.8	254401	2	AC126966	AC126966 Rattus no	C 154	17.6	70.4	78371	9	HSU66824	HSU66824 Human DNA
82	18.2	72.8	255066	2	AC103019	AC103019 Rattus no	C 155	17.6	70.4	79175	9	AC012516	AC012516 Homo sapi
83	18.2	72.8	255367	2	AC121743	AC121743 Rattus no	C 156	17.6	70.4	89139	9	EX000531	EX000531 Human DNA
84	18.2	72.8	268925	2	AC095911	AC095911 Rattus no	C 157	17.6	70.4	91940	9	AC112254	AC112254 Homo sapi
85	18.2	72.8	290885	2	AC128572	AC128572 Rattus no	C 158	17.6	70.4	102748	9	AC128708	AC128708 Homo sapi
86	18.2	72.8	295134	3	AE003582	AE003582 Drosophil	C 159	17.6	70.4	104423	2	HSB7824K2	HSB7824K2 Homo sapi
87	18.2	72.8	30141	9	HSEWSGAR	Y07848 Homo sapien	C 160	17.6	70.4	110000	2	AC111750	AC111750 Continuation (3 of
88	18.2	72.0	126312	9	AC000026	AC000026 Homo sapi	C 161	17.6	70.4	110000	2	LMFLCHR25	LMFLCHR25 Continuation (7 of
89	18.2	72.0	173029	9	AC002059	AC002059 Homo sapi	C 162	17.6	70.4	112061	9	BX247947	BX247947 Human DNA
90	18.2	72.0	202490	9	AC138066	AC138066 Pan trogl	C 163	17.6	70.4	112578	9	AL645927	AL645927 Human DNA
91	18.2	72.0	239803	2	AC113785	AC113785 Rattus no	C 164	17.6	70.4	113884	9	AL512442	AL512442 Human DNA
92	18.2	72.0	245121	2	AC126482	AC126482 Rattus no	C 165	17.6	70.4	114868	9	HS994B9	HS994B9 Human DNA
93	18.2	72.0	294286	2	AC137391	AC137391 Rattus no	C 166	17.6	70.4	117417	9	AL353704	AL353704 Human DNA
94	18.2	72.0	299784	2	AC137392	AC137392 Rattus no	C 167	17.6	70.4	118787	10	AL844486	AL844486 Mouse DNA
95	17.8	71.2	175	9	HSAPX7020	AC097020 Rattus no	C 168	17.6	70.4	127045	10	AL669969	AL669969 Mouse muscu
96	17.8	71.2	464	9	AF540541	U5250 Human paire	C 169	17.6	70.4	127219	2	AC023806	AC023806 Mus muscu
97	17.8	71.2	702	11	BV054540	AF540541 Pyxine fa	C 170	17.6	70.4	129573	2	AL671618	AL671618 Homo sapi
98	17.8	71.2	703	11	BV029896	BV054540 S212P6206	C 171	17.6	70.4	133836	9	AL662869	AL662869 Human DNA
99	17.8	71.2	35878	9	AP000549	BP002896 S212P6625	C 172	17.6	70.4	134514	9	AC025192	AC025192 Homo sapi
100	17.8	71.2	50592	9	AP000549	AC006107 Homo sapi	C 173	17.6	70.4	138203	1	AY310323	AY310323 Streptomy
101	17.8	71.2	76763	9	AL157705	AC006107 Homo sapi	C 174	17.6	70.4	140990	10	AC142257	AC142257 Mus muscu
102	17.8	71.2	110000	2	AC101727	AL157705 Human DNA	C 175	17.6	70.4	148851	9	HS155022	HS155022 Human DNA
103	17.8	71.2	113146	2	AC138440	AL101727 Mus muscu	C 176	17.6	70.4	148996	9	AC102953	AC102953 Homo sapi
104	17.8	71.2	116197	2	AC135614	AC138440 Tetraodon	C 177	17.6	70.4	149514	2	AP001937	AP001937 Homo sapi
105	17.8	71.2	127051	2	AC114895	AC135614 Pan trogl	C 178	17.6	70.4	154821	8	AP003709	AP003709 Oryza sat
106	17.8	71.2	128744	10	AL645803	AC114895 Tetraodon	C 179	17.6	70.4	157116	2	AC146358	AC146358 Otollemur
107	17.8	71.2	155749	2	AC135345	AL645803 Mouse DNA	C 180	17.6	70.4	157324	9	AC013643	AC013643 Homo sapi
108	17.8	71.2	161099	9	AC126332	AC135345 Homo sapi	C 181	17.6	70.4	163795	9	AP000356	AP000356 Homo sapi
109	17.8	71.2	165315	2	AC109304	AC126332 Homo sapi	C 182	17.6	70.4	166182	10	AC084391	AC084391 Mus muscu
110	17.8	71.2	167628	9	AC092570	AC109304 Mus muscu	C 183	17.6	70.4	168311	2	AC044841	AC044841 Homo sapi
111	17.8	71.2	172587	2	AC128231	AC092570 Homo sapi	C 184	17.6	70.4	171105	2	AC142367	AC142367 Rattus no
112	17.8	71.2	178473	9	HSJ745H19	AC128231 Rattus no	C 185	17.6	70.4	172022	2	AC145440	AC145440 Bos tauru
113	17.8	71.2	183179	2	AL133212	AL117380 Human DNA	C 186	17.6	70.4	172314	2	AC147360	AC147360 Mus muscu
114	17.8	71.2	187101	9	AL589862	AL133212 Homo sapi	C 187	17.6	70.4	172413	2	AC134388	AC134388 Pan trogl
115	17.8	71.2	187274	2	AC1391096	AL589862 Human DNA	C 188	17.6	70.4	174321	2	AC012286	AC012286 Homo sapi
116	17.8	71.2	188766	9	AC135329	AL391096 Homo sapi	C 189	17.6	70.4	175330	9	AC018457	AC018457 Homo sapi
117	17.8	71.2	193159	9	AC006946	AC135329 Homo sapi	C 190	17.6	70.4	175380	9	AF235104	AF235104 Homo sapi
118	17.8	71.2	195397	2	AC135718	AC006946 Homo sapi	C 191	17.6	70.4	176107	2	AP002393	AP002393 Homo sapi
119	17.8	71.2	196904	10	AC126447	AC135718 Homo sapi	C 192	17.6	70.4	176433	9	AC090246	AC090246 Homo sapi
120	17.8	71.2	200202	10	AC112154	AC126447 Mus muscu	C 193	17.6	70.4	177227	2	AC117009	AC117009 Rattus no
121	17.8	71.2	216645	2	AC119156	AC112154 Mus muscu	C 194	17.6	70.4	178802	2	AC140116	AC140116 Rattus no
122	17.8	71.2	218360	2	AC135708	AC119156 Mus muscu	C 195	17.6	70.4	180368	2	AC024276	AC024276 Homo sapi
123	17.8	71.2	227370	2	AC114867	AC135708 Rattus no	C 196	17.6	70.4	181047	2	AC016765	AC016765 Homo sapi
124	17.8	71.2	232191	2	AC096110	AC114867 Rattus no	C 197	17.6	70.4	181602	9	AC011208	AC011208 Homo sapi
125	17.8	71.2	233062	9	AC100787	AC096110 Rattus no	C 198	17.6	70.4	183008	2	AC019276	AC019276 Homo sapi
126	17.8	71.2	234094	2	AC130066	AC100787 Homo sapi	C 199	17.6	70.4	187341	2	AC024980	AC024980 Homo sapi
127	17.8	71.2	234950	2	AC094003	AC130066 Rattus no	C 200	17.6	70.4	187696	2	AC119757	AC119757 Rattus no
128	17.8	71.2	242416	2	AC133404	AC094003 Rattus no	C 201	17.6	70.4	188932	2	AC105176	AC105176 Mus muscu
129	17.8	71.2	246159	2	AC1098128	AC133404 Rattus no	C 202	17.6	70.4	189723	2	AC090952	AC090952 Homo sapi
130	17.8	71.2	246526	2	AC098128	AC1098128 Rattus no	C 203	17.6	70.4	191308	2	AC145062	AC145062 Pan trogl
131	17.8	71.2	246990	2	AC136585	AC098128 Rattus no	C 204	17.6	70.4	196712	5	BX323818	BX323818 Zebrafish
132	17.8	71.2	247875	2	AC137235	AC136585 Rattus no	C 205	17.6	70.4	199051	9	AC105224	AC105224 Homo sapi
133	17.8	71.2	248672	2	AC102953	AC137235 Rattus no	C 206	17.6	70.4	199993	10	AC102783	AC102783 Mus muscu
134	17.8	71.2	249541	2	AC106560	AC102953 Rattus no	C 207	17.6	70.4	200734	2	AC109272	AC109272 Mus muscu
135	17.8	71.2	250480	2	AC115644	AC106560 Rattus no	C 208	17.6	70.4	201757	10	AC061963	AC061963 Mus muscu
136	17.8	71.2	253461	2	AC127737	AC115644 Rattus no	C 209	17.6	70.4	202683	9	AC100865	AC100865 Homo sapi
137	17.8	71.2	255916	2	AC110316	AC127737 Rattus no	C 210	17.6	70.4	203782	10	AC110882	AC110882 Mus muscu
138	17.8	71.2	264902	2	AC126667	AC110316 Rattus no	C 211	17.6	70.4	205463	2	AC105308	AC105308 Pan trogl

C 212	17.6	70.4	205619	2	AC108275	Rattus no
C 213	17.6	70.4	208406	2	AC133405	Rattus no
C 214	17.6	70.4	209378	2	AC137262	Rattus no
C 215	17.6	70.4	209451	2	AC124942	Rattus no
C 216	17.6	70.4	216825	2	AC108908	Mus muscu
C 217	17.6	70.4	217595	10	AL805959	Mouse DNA
C 218	17.6	70.4	219315	2	AC119914	Mus muscu
C 219	17.6	70.4	219548	2	AC034265	Mus muscu
C 220	17.6	70.4	221154	2	AC111543	Rattus no
C 221	17.6	70.4	221968	2	AC115188	Rattus no
C 222	17.6	70.4	224973	2	AC118082	Rattus no
C 223	17.6	70.4	233042	10	AC132144	Mus muscu
C 224	17.6	70.4	234429	9	AF311103	Homo sapi
C 225	17.6	70.4	235978	2	AC094209	Rattus no
C 226	17.6	70.4	236760	2	AC132537	Rattus no
C 227	17.6	70.4	237644	2	AC095758	Rattus no
C 228	17.6	70.4	237913	2	AC094576	Rattus no
C 229	17.6	70.4	239486	2	AC094381	Rattus no
C 230	17.6	70.4	240657	2	AC111675	Rattus no
C 231	17.6	70.4	242507	2	AC134246	Mus muscu
C 232	17.6	70.4	242825	9	HSPEX	
C 233	17.6	70.4	244376	2	AC120968	Rattus no
C 234	17.6	70.4	246688	2	AC112282	Rattus no
C 235	17.6	70.4	247504	2	AC101851	Mus muscu
C 236	17.6	70.4	249206	2	AC145168	Mus muscu
C 237	17.6	70.4	252175	2	AC107539	Rattus no
C 238	17.6	70.4	252283	2	AC126977	Rattus no
C 239	17.6	70.4	252348	2	AC094064	Rattus no
C 240	17.6	70.4	253409	2	AC106926	Rattus no
C 241	17.6	70.4	253819	2	AC131329	Mus muscu
C 242	17.6	70.4	255470	2	AC106968	Rattus no
C 243	17.6	70.4	260164	2	AC098181	Rattus no
C 244	17.6	70.4	260589	2	AC126862	Rattus no
C 245	17.6	70.4	263069	2	AC103440	Rattus no
C 246	17.6	70.4	263388	2	AC094260	Rattus no
C 247	17.6	70.4	273663	2	AC133292	Rattus no
C 248	17.6	70.4	275060	2	AC135406	Rattus no
C 249	17.6	70.4	275924	2	AC116233	Rattus no
C 250	17.6	70.4	303438	1	AE017212	Geobacter
C 251	17.6	70.4	305225	2	AC130763	Rattus no
C 252	17.6	70.4	312905	2	AC094521	Rattus no
C 253	17.6	70.4	347800	1	AP000060	Aeropyrum
C 254	17.4	69.6	671	11	BV014309	S212P6736
C 255	17.4	69.6	36235	3	LMPL7836	
C 256	17.4	69.6	110000	2	AC109500_3	
C 257	17.4	69.6	110000	2	AC120598_2	
C 258	17.4	69.6	117000	9	AC069222	Homo sapi
C 259	17.4	69.6	124102	9	AC021148	Homo sapi
C 260	17.4	69.6	126589	9	AL512444	Human DNA
C 261	17.4	69.6	133289	9	AC012317	Homo sapi
C 262	17.4	69.6	148179	9	AP001831	Homo sapi
C 263	17.4	69.6	150522	9	AC008870	Homo sapi
C 264	17.4	69.6	153024	2	AC133073	Lemur cat
C 265	17.4	69.6	156604	2	AL390194	Homo sapi
C 266	17.4	69.6	158302	2	AC069481	Homo sapi
C 267	17.4	69.6	160212	2	AC009043	Homo sapi
C 268	17.4	69.6	163962	9	BS000233	Pan trogl
C 269	17.4	69.6	164736	9	AC105275	Homo sapi
C 270	17.4	69.6	166992	2	AC022883	Homo sapi
C 271	17.4	69.6	171553	2	AC013611	Homo sapi
C 272	17.4	69.6	178967	2	AC128772	Rattus no
C 273	17.4	69.6	183692	8	AP003230	Oryza sat
C 274	17.4	69.6	183988	2	AC012185	Homo sapi
C 275	17.4	69.6	188292	2	AC103360	Mus muscu
C 276	17.4	69.6	196086	10	AC104889	Mus muscu
C 277	17.4	69.6	197909	10	AL928719	Mouse DNA
C 278	17.4	69.6	204239	10	AC114819	Mus muscu
C 279	17.4	69.6	204693	2	AC131733	Mus muscu
C 280	17.4	69.6	211456	10	AL928926	Mouse DNA
C 281	17.4	69.6	217521	2	AC016047	Homo sapi
C 282	17.4	69.6	219253	2	AC110939	Rattus no
C 283	17.4	69.6	219892	2	AC127200	Rattus no
C 284	17.4	69.6	227141	2	AC109109	Rattus no
C 285	17.4	69.6	228894	2	AC133735	Rattus no
C 286	17.4	69.6	229350	2	AC107166	Rattus no
C 287	17.4	69.6	233220	2	AC098058	Rattus no
C 288	17.4	69.6	236716	2	AC098920	Rattus no
C 289	17.4	69.6	267088	2	AC129736	Rattus no
C 290	17.4	69.6	269863	2	AC131017	Rattus no
C 291	17.4	69.6	275605	2	AC123100	Rattus no
C 292	17.4	69.6	278166	2	AC094117	Rattus no
C 293	17.4	69.6	288483	2	AL627109	Homo sapi
C 294	17.4	69.6	340000	9	AP001748	Homo sapi
C 295	17.2	68.8	461	9	AY027915S2	
C 296	17.2	68.8	529	11	BV070169	
C 297	17.2	68.8	640	11	G96635	
C 298	17.2	68.8	1126	4	PIGGR0AI	
C 299	17.2	68.8	2000	6	AX656630	Sequence
C 300	17.2	68.8	3014	4	AY267904	Sus scrofa
RESULT 1						
AE006916						
LOCUS						
DEFINITION	Myobacterium tuberculosis CDC1551, section 2 of 280 of the complete genome.					
ACCESSION	AE006916					
VERSION	AE006916.1					
KEYWORDS	Myobacterium tuberculosis CDC1551					
SOURCE	Myobacterium tuberculosis CDC1551					
ORGANISM	Myobacterium tuberculosis CDC1551					
REFERENCE	1 (bases 1 to 15619)					
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.					
TITLE	Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 15619)					
AUTHORS	Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.					
TITLE	Direct Submission					
JOURNAL	Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA					
FEATURES	source					
gene						
CDS						

AC133735 Rattus no  
AC107166 Rattus no  
AC098058 Rattus no  
AC098920 Rattus no  
AC129736 Rattus no  
AC131017 Rattus no  
AC123100 Rattus no  
AC094117 Rattus no  
AL627109 Homo sapi  
AP001748 Homo sapi  
AY027915 Homo sapi  
BV070169 S212P6630  
G96635 S208P6068RF  
L07317 Pig G prote  
AX656630 Sequence  
AY267904 Sus scrofa

15619 bp DNA linear BCT 27-APR-2001

Myobacterium tuberculosis CDC1551, section 2 of 280 of the complete genome.

AE006916 AE0000516  
AE006916.1 GI:13879055

Myobacterium tuberculosis CDC1551  
Myobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;  
Corynebacteriaceae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

1 (bases 1 to 15619)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.

Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains

Unpublished  
2 (bases 1 to 15619)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.

Direct Submission  
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers  
1. 15619  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/notes="clinical strain"  
78. .866  
/gene="MT0015"  
78. .866  
/gene="MT0015"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK44237.1"  
/db\_xref="GI:13879056"  
/translation="MELTHTPCPENGEMTIDRRSARFSPVLCVLAAGLIAATHG  
VSGGTIRSDAPRLVDLYRQAQSVNRLATEREALTTRIDSVHGSDVDTALAAOER  
SAGLGAAMNPFVHGGLVVLQDAORDANGRPDRASFDLIVHQQDLEAVLNALWN  
AGAEATQMDQRIIAIARCVNTLLNGRTYSPPTTAAIGADAAAMQALAAAPLV



```
gene
CDS
TLKQVYVRFGLGYREEVHPDLQIVGYADPVRMHFAQAGPLDY"
/903..1601
/gene="MT0016"
/903..1601
/gene="MT0016"
/notes="similar to PID:1262359; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamine amidotransferase, class I"
/protein_id="AAK44238.1"
/db_xref="GI:13879057"
/transl_table=11
/translation="MRILVVDNYSFVFNLYQVLCGLGIEAEVWRNDRHLSDEAAVA
GQFDVLLSPGCTPERAGASYSIVHACAAHTPLGLGCHQALGVAFATVRAPE
LHGKTSVFNHNVGLQGLDPPTATRYHSLTILPKSLPAVLAVTARTSSGVIMAVQ
HTGLPIHGQVFPESILTEGGHRIILANLWTCGWTQDDTLVRRLENEVLIALSPHEPT
STASAGEATGRTSA"
/complement(1579..3459)
/gene="MT0017"
/complement(1579..3459)
/gene="MT0017"
/notes="similar to PID:2344821; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine/threonine protein kinase"
/protein_id="AAK44239.1"
/db_xref="GI:13873058"
/translation="MTTSHLSDRVELGELGILFGGSEVHLARDLRHRDVAVKVLRA
DIARDFSVLRRAQAALNPAIVAVIDGSAETPAGLPPIYVYVDGVTLRD
IVHTGPTMKAEIVADACQALNQHONGIIRDAVKPAMISATNAVXMDFGIA
RAIADSGVOTTAIVTAQVLSPEQAGSDVARSVSLGCVLYELVTEGPEPTFV
DSFVSVAHVREDIPPTASHEGLSADLDAVLKALKNPENRYQTAAENRADLVTV
HNGEPEAPKVTDAERTSLSSAAGNLSPRTDPLQDLDLDRDSIGSGVGRWA
VNAVEALVTVVITAINPFGGITRDVQVQVDRGSSADAIALQNRGFKIRLTAGGEG
TIPDRVIGTDAANTSAGDEIVTVSTGPEQREIPDVSTLTVAEAVKLTAAEGG
RKQNSPTSELVKVIGTNPENQTSAINTVNIIIVGSGPAIKDIEDVAGQVDDVA
QKLNAYGTFKFSQSVSPRAGVETGTPAGTVPVDSVIELQVSKGNQFVMPDL
SGMFWYDAEPRLALGWTGMLDKGADVAGGSQHNRYVYQNPAGTGVNRDGIITLRF
GQ"
/complement(3456..4751)
/gene="MT0018"
/complement(3456..4751)
/notes="similar to PID:2344822; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="serine/threonine protein kinase"
/protein_id="AAK44240.1"
/db_xref="GI:13879059"
/translation="NSPRGVTLGSRYLQRLIATGCGQVWEAVDNELRRVAVKVL
KSEFSDPEFIRFRAERTTAMLNHPGIASVHDYGESQMGEGRTAYLVNMLVNGEP
LNSVLRKTRLSRLALDMEQTLGALQIAAAGLVHRDVPKNILITPTQGVKITDF
GKAKVADAPVOTGMWMTAQVIAPEQALGHDASPDVSLGVVGYEAVSGKPPFA
GKATVAMKHIKEPPPLPDLPNVRELIEITLVKNPAMEYRSGGPFADAVAARA
GRPPRPSPTPPGRAAPAIISGTTAVANSAGRTAASRSPRATGHPHPPRTFS
SGRALLWAAGVLGALIIIVLVKAPGNSPQATPTPTVTITGNPNPASNTGGTDA
SPRLNNWGETRHSGLQSVVPPPHGRASLARYEIAQ"
/complement(4748..6223)
/gene="MT0019"
/complement(4748..6223)
/gene="MT0019"
/notes="similar to GP:5139605; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein, putative"
/protein_id="AAK44241.1"
/db_xref="GI:13879060"
/translation="WNASLRRIISVTMALIVLLINATMTQVFTADGLRASPQRVL
LDEYSRQRCQITAGGQLLAYSVATGRRFRFLVYENPSPYAPVTCFYSLSRYSTALER
ASDPIILNGDRRLFRRLADFFTGDRDPRGNGVDITINPIQAGHDAMQGCYGFCKG
AVVALPSTGKILALVSSPSYDFNLIASHNPEVQAQWQRLGDNFASPLTNRAISETY
PGSTFKVITTAALAAAGATETQLTAAPTIPLPGSTAOLENYGAGCDEPTVSLRE
AFVKSNTAFVOLGRTGADALRSMAARGLDSPRPPTPLOVAESTVGPIPSAALGM
TSIGQDVALTPLANEAIAATANGGITNRPYLVSLGKGLDLANISTTVGYQRRAVS
POVAAKLTMLMGAERKVAQKGAIPGVQASKITGAETHGTDPRHIPPHAWYIAFAPAQ
AFKVAVALVENGADRLSATGALLAIPICRAVIAEALQGEF"
/complement(6220..7629)
/gene="MT0020"
/complement(6220..7629)
/gene="MT0020"
/notes="similar to GB:DI0483 SP:PI6457 GB:M30807 PID:146039
PID:216503; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsW"
/protein_id="AAK44242.1"
/db_xref="GI:13879061"
/translation="MTTRLOAPVAVTPPTTRNAELLLLCFAAVITPAALLVQANO
DQGVPMDLTSYGLAFTLFGSAHLAIRRFAPYTDPLLPVALLNGLGLVMHRLDLV
DNEIGEHRRPSANQOMLTLGVAAFPALVTFKQHRQARYGICGLAGLVFLAVPA
LLPALSEQNGAKIWLRFQFSIQPAEFKILLIFFSAVLAKRGLFSAGKLLGM
TLPRDLAPLAAAVISQVMVFEKDLGSLILLYTSFLVVLATRFSTVWVIGLTL
FRAGTIVAYFIPEHVELRVQTLDPADPDGTQIVQVLSFSGFATGGIFGTGLNGQP
DTVPAASTDFIITAAEGEELGLTALMLYTIIVIRGLRTAIAIRDSFGKLLAAGLS
STIAIQFIIVGVGTSLPLTGLTTFWMSYGGSSLLANLAILALARIHSGARRPLRT
RPRNKSPIITAAAGTVEIRV"
/complement(7626..9161)
/gene="MT0021"
/complement(7626..9161)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44243.1"
/db_xref="GI:13879062"
/translation="MTLVLYAARSRLVRANEDSVYAGARLLALADCMGHAAGE
VASQVLAALHLDDEPGRDILLKLDAAVRAGNSAIQAQVMEPDEMGMTTLTAIL
FAGNRGLVHIGSRGYLLRDGELTQITKDDTFVQTLVDEGRITPEEAHSHPQSLIM
RAITGHEVETLTREARADRYLLCSGLSDPSVDETILEALQIPEVAESARLIEL
ALRGGPDNVTVVADVVDYDQGTQPILAGAVSGDDQTLPTNPAAGSAISQKKE
IVKRVPPQADTFSPRPSWRRLAFVAVTVLMTAGLLGIRAIIRSNVYVAYAGSVS
IVRGIQGLLGMSLHOPYLMGCLSPNELSQISYSGSGGLDCHLMKLEDLRPPRAQ
VRAGLPAGTLDDAIGOLRELAANSLPPCAPRATSPGCRPAPTSETTEPNTSSP
ASPSPTASAPGTGTPAIPTSASPAAPASPTPMPPTSSPTMAALPPPPPQPGIDCR
AAA"
/complement(9259..9726)
/gene="MT0022"
/complement(9259..9726)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44244.1"
/db_xref="GI:13879063"
/translation="MQGLVLQTRAGFLMLLWVFIWSVLRIKLTIDYAFPTGVMRRRG
LAIRGTLGARQRHARHARVAVTEGALTGTARITLSEQVLIGRADSDTLVLDVAST
RHARLSMRSEWVEDLGSNTCTYLDRAKVTAVRPICTPVRIGKTAELRP"
/complement(9850..11415)
/gene="MT0023"
/complement(9850..11415)
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44245.1"
/db_xref="GI:13879064"
/translation="MGSKRLVQRVERKLEQTVGDFAFAIFGSIIVQVEALLRREA
ADIGISQGNRLAPNEYIITGLVHDFEKLGDAPPELKSTGFARLDADYIQEQGWQTYG
DVVVRTEOSNLHTGQFRAGITVNPDPVTHPPVDCARQSNHAFAGEPVSMSNS
SYRGGQGRDPDEYIDRVARPOEDPRGGPDQGGSDPRGGYPPETGTPGYPQGPYPRP
```





FT BACTERIA. THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY. THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION ONCE IT IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-INDEPENDENT) ALONG DUPLEX DNA [CATALYTIC ACTIVITY: N deoxynucleoside triphosphate + N diphosphate + {DNA}N]."

FT /gene="dnaN"

FT /locus\_tag="Rv0002"

FT /product="DNA POLYMERASE III (BETA CHAIN) DNAN (DNA NUCLEOTIDYLTRANSFERASE)"

FT /protein\_id="CAAL6239.1"

FT /translation="MDAATTRVGLTDLFRLLRESFADAVSWAKNLPARPAVFLVSGV LITGDSGLTISGFVYEAQVAGAEIVSPGSLVSGRLISDITRALPNKPFVDVHVEG NRVALTCGNARSLPTMVEDVPTUPTLPETGLLPAELFAEAIQVAVIAAGRDDTLPM LTGIVEILGFVNLAAIDRSLAVRELKMSASPDIEAANLVPAKTLAEAKAGIGS DVLISLGTGFGVCKGLGIGSGNGKRSITRLIDREFPKFRQLLTETHTAVATMDVABLI EAKILVALVDAGAGVQRMEFAGDSVRLSAGADDVGRAEEDLVVDYAGEPLTIANPYLL TDGLSLRSERVSFGFTTAGPALLRPVSGDDRPVAGLNGNGPFPFVASTDYVLLMPVR LFG"

FT CDS

FT 3280..4437

FT /evidence=EXPERIMENTAL

FT /note="Rv0003, (MTCY10H4.01), len: 385 aa, recF, DNA replication and repair protein (see citations below), equivalent to others Mycobacterium DNA replication and repair proteins e.g. NP\_301131.1|NC\_002677 from Mycobacterium leprae (385 aa); Q9L7L5|REFC\_MYCPA from Mycobacterium avium subsp. paratuberculosis (385 aa); P50916|REFC\_MYCSM from Mycobacterium smegmatis (384 aa); etc. Also highly similar to others e.g. P36176|REFC\_STRCO DNA REPLICATION AND REPAIR PROTEIN from Streptomyces coelicolor (373 aa); NP\_440892.1|NC\_000911 from Synecocystis sp. strain PCC 6803 (384 aa); NP\_469352.1|NC\_003212 from Listeria innocua (370 aa); etc. Contains PS00017 ATP/GTP-binding site motif A (P-loop), PS00617 RecF protein signature 1, and PS00618 RecF protein signature 2. BELONGS TO THE RECF FAMILY."

FT /transl\_table=1

FT /function="THE RECF PROTEIN IS INVOLVED IN DNA METABOLISM AND RECOMBINATION; IT IS REQUIRED FOR DNA REPLICATION AND NORMAL SOS INDUCIBILITY. RECF BINDS PREFERENTIALLY TO SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS TO BIND ATP."

FT /gene="recF"

FT /locus\_tag="Rv0003"

FT /product="DNA REPLICATION AND REPAIR PROTEIN RECF (SINGLE-STRAND DNA BINDING PROTEIN)"

FT /protein\_id="CAB02424.1"

FT /translation="MYVRLGLRDPFRSWACVDLELHPGRTVFVGPNGYKTNLIEALWY SFTLGHRSVADPLIRVGTDRVAVSTIVNDGRECAVDLEIATGRVNAKRLNRSSVRS TRDVGVLRVAFEDLGLVGRDPAARRYLDDLAIVRRPAIAVRAEYERVLVORTA LLKSVPGARYGRDGVFTLEWDSRLAEGELVAARIIDLVNOLAPEVKAYKOLLAPE SRASIGTRASNDVTGPEQSDIDQLLAALLAARDAELERGLGVGPRDDLI LRLGDPKAGFASGWSLVAUQLRLVQDGEPEVLLDDVFALDDVMRRALA TAESAQVLTVAVLEIDIPAGWDARRVHIDVADDTGSMGVLP"

FT misc\_feature

FT 3367..3390

FT /note="PS00017 ATP/GTP-binding site motif A"

FT 3634..3690

FT /note="PS00617 RecF protein signature 1"

FT 4243..4296

FT /note="PS00618 RecF protein signature 2"

FT 4434..4497

FT /evidence=EXPERIMENTAL

FT /note="Rv0004, (MTCY10H4.02), len: 187 aa. Conserved hypothetical protein (see Salazar et al., 1996), highly similar, but longer 21 aa in N-terminus, to AAF33696.1|AF222789 unknown protein from Mycobacterium avium subsp. paratuberculosis (166 aa); and highly similar to NP\_301132.1|NC\_002677 conserved hypothetical protein from Mycobacterium leprae (189 aa); S70990 hypothetical protein from Mycobacterium smegmatis (194 aa). Also highly similar, except in N-terminal part, to

Query Match 100.0%; Score 25; DB 15; Length 341957;  
Best Local Similarity 100.0%; Pred. No. 0.78;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGCACATAGC 25  
|||||  
Db 24426 CCGGAGGGCCAGAGGCACATAGC 24450  
|||||

RESULT 3  
LOCUS BX248334 343050 bp DNA linear BCT 11-JUN-2003  
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 1/14.  
ACCESSION BX248334 BX248333  
VERSION BX248334.1 GI:31616762  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis subsp. bovis AF2122/97  
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1  
AUTHORS Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthey, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P. R., Parkhill, J., Barrrell, B. G., Cole, S. T., Gordon, S. V. and Hewinson, G.  
TITLE The complete genome sequence of Mycobacterium bovis  
JOURNAL Online Publication  
REMARK PNAS 10.1073/pnas.1130426100 ( Microbiology )  
REFERENCE 2 (bases 1 to 343050)  
AUTHORS Garnier, T.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex 15, France. e-mail: tgarnier@pasteur.fr Submitted on behalf of the Mycobacterium bovis sequencing teams, TB Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. P4 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
UNITE de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES  
source  
1..343050  
/organism="Mycobacterium bovis subsp. bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
1..1524  
/gene="dnaA"  
/locus\_tag="Mb0001"  
1..1524  
/locus\_tag="dnaA"  
/locus\_tag="Mb0001"  
/note="Mb0001, dnaA, len: 507 aa. Equivalent to Rv0001, len: 507 aa, from Mycobacterium tuberculosis strain H37Rv, (99.6% identity in 507 aa overlap). dnaA, chromosomal replication initiator protein (see citations below), equivalent to other Mycobacterial CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g. P46388|DNAA\_MYCPA from Mycobacterium leprae (502 aa); Q9L7L7|DNAA\_MYCPA from Mycobacterium paratuberculosis (509 aa); P49990|DNAA\_MYCAV from Mycobacterium avium (508 aa); P49992|DNAA\_MYCSM from Mycobacterium smegmatis (504 aa); etc. Also highly similar to others except in N-terminus e.g. Q9ZH75|DNAA\_STRCH CHROMOSOMAL REPLICATION INITIATOR PROTEIN from Streptomyces chrysomallus (624 aa); Q9ZH76|DNAA\_STRSE from Streptomyces reticuli (643 aa); DNAA\_ECOLI|P03004|B3702 chromosomal replication initiator protein from Escherichia coli strain K12 (467 aa), FASTA scores: opt: 986, E(): 0, (43.2% identity in 389 aa overlap); etc. Contains PS00017 ATP/GTP-binding site motif A (P-loop) and PS01008 DnaA protein signature. BELONGS TO THE DNAA FAMILY. Note that

```

6903 (384 aa); NP_469352.1|NC_003212 from Listeria innocua
(370 aa); etc. Contains PS00017 ATP/GTP-binding site motif
A (p-loop), PS0617 RecF protein signature 1, and PS0618
RecF protein signature 2. BELONGS TO THE REC_F FAMILY."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="DNA REPLICATION AND REPAIR PROTEIN REC_F
(SINGLE-STRAND DNA BINDING PROTEIN)"
/protein_id="CAD92865.1"
/db_xref="GI:31616765"
/translation="MYVHRGLDPRFSWACVDLELHFQRTVFGPNGYGKTNILEALM
RSTTGVVRSADLVLPEDGLHVRGDIPDRRLDDCAIVRAIPAARAEYKRVVRQI
RTALLSGPVARVGROGVDFDTLVWDSSRLAEHGAEILVAARILDVLNOLAPEVKAYQL
LAPESASIGQPVASHNDVTGPSESDTDRIQLLAARLLAAALAREDALESIGCVKLQPH
RDDELIIRLGDPKANGFASGHQSADTRILAAATOLLRVDCGEFVLIDSDVFDELDM
RRALATAAESAEQVLTVAVFLDIPAGWDARRVHIIDRADDDTGSMNSVLLP"
4434..4997
/locus_tag="Mb0004"
/gene
CDS
4434..4997
/locus_tag="Mb0004"
/locus_tag="Mb0004"
/note="Mb0004, -, len: 187 aa. Equivalent to Rv0004, len:
187 aa, from Mycobacterium tuberculosis strain H37RV,
(99.5% identity in 187 aa overlap). Conserved hypothetical
protein (see citation below), highly similar, but longer
21 aa in N-terminus, to AAF3696.1|AF22789 unknown
protein from Mycobacterium avium subsp. paratuberculosis
(166 aa); and highly similar to NP_301132.1|NC_002677
conserved hypothetical protein from Mycobacterium leprae
(189 aa); S7090 hypothetical protein from Mycobacterium
smegmatis (194 aa). Also similar to in C-terminus to
C-terminal part of P35925|YREG_STRCO HYPOTHEICAL 19.8 KDa
PROTEIN (IN REC-F GYRB INTERGENIC REGION) from Streptomyces
coelicolor (190 aa), FASTA scores: opt: 404, E(): 3.9e-18,
(40.7% identity in 189 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD92866.1"
/db_xref="GI:31616766"
/translation="MTGSVDPRDNGERLMKSPCLDVRRITDLAEAAAARGQDAG
RGKVASVASGRVAGRNRSGSGPDPIROPPLGKAARELAKKRGSVRAEVMVLGVO
SAVVGHQIAEHARPTLRHLDGVLVSIAESTAWATQRIMQAOLAKIAAAVGNVDVRS
KITGPAAFPWRKGPRHIAQGPRDTYG"
5123..7267
/gene="gyrB"
/locus_tag="Mb0005"
5123..7267
/gene="gyrB"
/locus_tag="Mb0005"
/EC_number="5.99.1.3"
/note="Mb0005, gyrB, len: 714 aa. Equivalent to Rv0005,
len: 714 aa, from Mycobacterium tuberculosis strain H37RV,
(99.9% identity in 714 aa overlap). gyrB, DNA gyrase
subunit B (EC 5.99.1.3) (see citations below), equivalent,
except in N-terminus, to other Mycobacterial DNA GYRASES
SUBUNIT B e.g. T10005 from Mycobacterium leprae (697 aa);
Q9UL713|GYRB_MYCPA from Mycobacterium avium subsp.
paratuberculosis (677 aa) (has its N-terminus shorter);
P48355|GYRB_MVCSM from Mycobacterium smegmatis (675 aa);
etc. Also highly similar to others e.g. T10969 from
Streptomyces coelicolor (686 aa); P50075|GYBS_STRSH from
Query Match 100.0%; Score 25; DB 1; Length 343050;
Best Local Similarity 100.0%; Pred.No.0.78;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 CCCGGAGGCCAGAGGCACATAGC 25
|||||||
24425 CCCGGAGGCCAGAGGCACATAGC 24449

```

```

RESULT 4
MSGY219/c
LOCUS       38721 bp      DNA      linear      BCT 10-DEC-1996
DEFINITION  Mycobacterium tuberculosis sequence from clone y219.
ACCESSION   AD000013
VERSION     AD000013.1  GI:1717736
KEYWORDS    Mycobacterium tuberculosis
SOURCE      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
            tuberculosis complex.
REFERENCE   1 (bases 1 to 38721)
            Du, L.,
            Direct Submission
            Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
            Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT     GSDB:S:1004710
FEATURES             Location/Qualifiers
     source          1..38721
                     /organism="Mycobacterium tuberculosis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:1773"
                     /clone="y219"

ORIGIN
Query Match      96.0%; Score 24; DB 1; Length 38721;
Best Local Similarity 100.0%; Pred. NO. 2.7;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2  CCGAGGGCCAGAGGCACATAGC 25
      |||||
DB  37274 CCGAGGGCCAGAGGCACATAGC 37251

RESULT 5
AC016557
LOCUS       60740 bp      DNA      linear      HTG 25-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-374J18, *** SEQUENCING IN
ACCESSION   AC016557
VERSION     AC016557.8  GI:15290343
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 60740)
            DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS     DOE Joint Genome Institute
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 60740)
            DOE Joint Genome Institute.
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Aug 25, 2001 this sequence version replaced gi:13699576.
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 8415: contig of 8415 bp in length
            * 8416 8515: gap of unknown length
            * 8516 60740: contig of 52225 bp in length.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----

```

```

Project Information
Center Project Name: 329028
Center clone name: CTC-374J18
-----
Summary Statistics
Consensus quality: 58774 bases at least Q40
Consensus quality: 58860 bases at least Q20
Consensus quality: 58898 bases at least Q20
Estimated insert size: 108000; agarose-fp estimation
Estimated insert size: 60640; sum-of-contigs estimation
Quality coverage: 0.55 in Q20 bases; agarose-fp
estimation
Quality coverage: 0.97 in Q20 bases; sum-of-contigs
estimation.
FEATURES             Location/Qualifiers
     source          1..60740
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-374J18"

ORIGIN
Query Match      80.8%; Score 20.2; DB 2; Length 60740;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY  1  CCGGAGGGCCAGAGGCACATAGC 25
      |||||
DB  13110 CCGGATGGCCAGCGGCACAGAGC 13134

RESULT 6
AC008544/c
LOCUS       168008 bp      DNA      linear      PRI 05-APR-2002
DEFINITION  Homo sapiens chromosome 5 clone CTC-500G13, complete sequence.
ACCESSION   AC008544
VERSION     AC008544.7  GI:20043093
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168008)
            DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS     DOE Joint Genome Institute
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 168008)
            DOE Joint Genome Institute.
AUTHORS     DOE Joint Genome Institute
TITLE       Direct Submission
JOURNAL     Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            3 (bases 1 to 168008)
            DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS     DOE Joint Genome Institute
TITLE       Direct Submission
JOURNAL     Submitted (05-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            4 (bases 1 to 168008)
            DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS     DOE Joint Genome Institute
TITLE       Direct Submission
JOURNAL     Submitted (05-APR-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            On Apr 5, 2002 this sequence version replaced gi:18497066.
COMMENT     Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www.shgc.stanford.edu
            Quality: Phrap Quality >=40 99.6% of Sequence;
            Estimated Total Number of Errors is 0.3.
            -----Location/Qualifiers
            1..168008
            /organism="Homo sapiens"
            /mol_type="genomic DNA"

```

db_xref="taxon:9606"	80.8%;	Score 20.2;	DB 9;	Length 168008;
Query Match	80.8%;	Score 20.2;	DB 9;	Length 168008;
Best Local Similarity	88.0%;	Pred. No. 1.1e+02;		
Matches 22;	Conservative 0;	Mismatches 3;	Indels 0;	Gaps 0;
ORIGIN				
1	CCCCGAGGGCCAGAGGGCCACATAGC 25			
33928	CCCCGATGGCCAGCGGCACACAGC 33904			
Db				
RESULT 7				
AC034202	171004 bp	DNA	linear	HTG 20-APR-2001
LOCUS	28	unordered pieces.		
DEFINITION	28	unordered pieces.		
ACCESSION	AC034202			
VERSION	AC034202.4	GI:13699719		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.			
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.			
ORGANISM	Homo sapiens			
REFERENCE				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	1 (bases 1 to 171004)			
JOURNAL	DOE Joint Genome Institute.			
REFERENCE	2 (bases 1 to 171004)			
AUTHORS	DOE Joint Genome Institute.			
TITLE	Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint			
JOURNAL	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
COMMENT	On Apr 20, 2001 this sequence version replaced gi:7712062.			
-----Genome Center				
Center: Joint Genome Institute				
Center Code: JGI				
Web site: http://www.jgi.doe.gov				
-----				
Project Information				
Center Project Name: 131388				
Center clone name: CIT978SKB_174P8				
-----				
Summary Statistics				
Consensus quality: 144087 bases at least Q40				
Consensus quality: 156868 bases at least Q30				
Consensus quality: 161298 bases at least Q20				
Estimated insert size: 141000; pulse field gel estimation				
Estimated insert size: 168304; sum-of-contigs estimation				
Quality coverage: 6.14 in Q20 bases; pulse field gel estimation				
Quality coverage: 5.14 in Q20 bases; sum-of-contigs estimation				
* NOTE: This is a 'working draft' sequence. It currently				
* consists of 28 contigs. The true order of the pieces				
* is not known and their order in this sequence record is				
* arbitrary. Gaps between the contigs are represented as				
* runs of N, but the exact sizes of the gaps are unknown.				
* This record will be updated with the finished sequence				
* as soon as it is available and the accession number will				
* be preserved.				
1	1214: contig of 1214 bp in length			
1215	1314: gap of unknown length			
1315	2426: contig of 1112 bp in length			
2427	2526: gap of unknown length			
2527	3572: contig of 1046 bp in length			
3573	3672: gap of unknown length			
3673	5017: contig of 1345 bp in length			
5018	5117: gap of unknown length			
5118	6269: contig of 1152 bp in length			
6270	6369: gap of unknown length			
6370	8139: contig of 1770 bp in length			
8140	8239: gap of unknown length			
8240	10204: contig of 1965 bp in length			
10205	10304: gap of unknown length			
10305	11517: contig of 1213 bp in length			
11518	11617: gap of unknown length			
11618	15037: contig of 3420 bp in length			
15038	15137: gap of unknown length			
15139	17640: contig of 2503 bp in length			
17641	17740: gap of unknown length			
17741	19599: contig of 1859 bp in length			
19600	19699: gap of unknown length			
19700	22850: contig of 3151 bp in length			
22851	22950: gap of unknown length			
22951	25617: contig of 2667 bp in length			
25618	25717: gap of unknown length			
25718	28276: contig of 2559 bp in length			
28277	28376: gap of unknown length			
28377	31321: contig of 2945 bp in length			
31322	31421: gap of unknown length			
31423	34125: contig of 2704 bp in length			
34126	34225: gap of unknown length			
34226	37463: contig of 3138 bp in length		</	



Center Project Name: 1705165		Center clone name: RPCI-11_1069G21	
-----			
Summary Statistics			
Consensus quality: 187962 bases at least Q40			
Consensus quality: 193144 bases at least Q30			
Consensus quality: 195917 bases at least Q20			
Estimated insert size: 184000; agarose-fp estimation			
Estimated insert size: 198268; sum-of-contigs estimation			
Quality coverage: 6.67 in Q20 bases; agarose-fp estimation			
Quality coverage: 6.19 in Q20 bases; sum-of-contigs estimation.			
* NOTE: This is a 'working draft' sequence. It currently			
* consists of 16 contigs. The true order of the pieces			
* is not known and their order in this sequence record is			
* arbitrary. Gaps between the contigs are represented as			
* runs of N, but the exact sizes of the gaps are unknown.			
* This record will be updated with the finished sequence			
* as soon as it is available and the accession number will			
* be preserved.			
* 1 1365: contig of 1365 bp in length			
* 1366 1465: gap of unknown length			
* 1466 2896: contig of 1431 bp in length			
* 2897 2996: gap of unknown length			
* 2997 4621: contig of 1625 bp in length			
* 4622 4721: gap of unknown length			
* 4722 7658: contig of 2937 bp in length			
* 7659 7758: gap of unknown length			
* 7759 11553: contig of 3795 bp in length			
* 11554 11653: gap of unknown length			
* 11654 14768: contig of 3115 bp in length			
* 14769 14868: gap of unknown length			
* 14869 18046: contig of 3178 bp in length			
* 18047 18146: gap of unknown length			
* 18147 25440: contig of 7294 bp in length			
* 25441 25540: gap of unknown length			
* 25541 30807: contig of 5267 bp in length			
* 30808 30907: gap of unknown length			
* 30908 38615: contig of 7708 bp in length			
* 38616 38715: gap of unknown length			
* 38716 50105: contig of 11390 bp in length			
* 50106 50205: gap of unknown length			
* 50206 74602: contig of 24397 bp in length			
* 74603 74702: gap of unknown length			
* 74703 99188: contig of 24486 bp in length			
* 99189 99288: gap of unknown length			
* 99289 128021: contig of 28733 bp in length			
* 128022 128121: gap of unknown length			
* 128122 160842: contig of 32721 bp in length			
* 160843 160942: gap of unknown length			
* 160943 199768: contig of 38826 bp in length.			
Location/Qualifiers			
1..199768			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/chromosome="5"			
/clone="RP11-1069G21"			
/clone_lib="RPCI human BAC library 11"			
ORIGIN			
Query Match 80.8%; Score 20.2; DB 2; Length 199768;			
Best Local Similarity 88.0%; Pred. No. 1.1e+02;			
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY 1 CCCGAGGGCCAGAGGGCACATAGC 25			
DB 101360 CCCGATGCCAGCGGGCACAGAGC 101336			
RESULT 10			
AC136594 199768 bp DNA linear HTG 06-NOV-2002			
LOCUS Homo sapiens chromosome 5 clone RP11-1069G21, WORKING DRAFT			
DEFINITION SEQUENCE, 16 unordered pieces.			
ACCESSION AC136594.1			
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.			
KEYWORDS Homo sapiens (human)			
SOURCE Homo sapiens			
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 199768)			
DOE Joint Genome Institute.			
Sequencing of Human Chromosome 5			
Unpublished			
2 (bases 1 to 199768)			
DOE Joint Genome Institute.			
Direct Submission			
Submitted (06-NOV-2002) Production Sequencing Facility, DOE Joint			
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
-----Genome Center			
Center: Joint Genome Institute			
Center Code: JGI			
Web site: http://www.jgi.doe.gov			
-----			
Project Information			
AC136596 220632 bp DNA linear HTG 06-NOV-2002			
LOCUS Homo sapiens chromosome 5 clone RP11-1070F6, WORKING DRAFT			
DEFINITION			







REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

MO 63108, USA  
 6 (Bases 1 to 104926)  
 Waterston,R.  
 Direct Submission  
 Submitted (29-MAY-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On May 3, 2002 this sequence version replaced gi:20304048.

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0103H17  
 -----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oseegawa K., Woon, P.Y., Zhao, B., Frengen, E., Taten, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP11-83A24, 2000 bp overlap; the clone sequenced to the right is RP11-308D13, 2000 bp overlap.  
 Actual start of this clone is at base position 113772 of RP11-83A24; actual end is at base position 61170 of RP11-308D13.

Data from AC112141 was used to finish AC114743.

#### FEATURES

Source	
1. .104926	Location/Qualifiers
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="4"	
/map="4"	
/clone="RP11-103H17"	
/clone.lib="RPCI-11"	
3. .140	
/rpt_family="L1"	
141. .447	
/rpt_family="MER2_type"	
448. .920	
/rpt_family="L1"	
1371. 1700	
/rpt_family="L1"	
1701. .1987	
/rpt_family="Alu"	
1988. .3303	
/rpt_family="L1"	
3304. .3594	
/rpt_family="Alu"	
3619. .4074	
/rpt_family="MALR"	
4170. .4255	
/rpt_family="L1"	
4256. .4339	
/rpt_family="MER2_type"	
4330. .4407	
/rpt_family="MER2_type"	
4408. .4486	
/rpt_family="MER2_type"	
4487. .4589	
/rpt_family="L1"	
5080. .5526	
/rpt_family="ERV1"	
5527. .5685	
/rpt_family="ERV1"	
5686. .5979	
/rpt_family="Alu"	
5980. .7106	
/rpt_family="ERV1"	
10347. .10654	
/rpt_family="Alu"	
10659. .11120	
/rpt_family="ERV1"	
11121. .11405	
/rpt_family="Alu"	
11477. .11638	
/rpt_family="(TG)n"	
11791. .11846	
/rpt_family="Alu"	
11847. .12060	
/rpt_family="ERV1"	
12061. .12444	
/rpt_family="ERV1"	
12445. .12464	
/rpt_family="(TA)n"	
12465. .12504	
/rpt_family="ERV1"	
13184. .13380	
/rpt_family="MER1_type"	
13583. .13886	
/rpt_family="Alu"	
13894. .14049	
/rpt_family="Alu"	
14052. .14361	
/rpt_family="Alu"	
15977. .17226	
/note="CpG island (%GC=67.1, o/e=0.80, #CpGs=120)"	
17300. .17424	
/rpt_family="MER1_type"	
18508. .18589	
/rpt_family="MIR"	
18590. .18889	
/rpt_family="Achoho"	
18690. .18808	
/rpt_family="MIR"	
19959. .19980	
/rpt_family="AT-rich"	
20365. .20515	
/rpt_family="MER1_type"	
20834. .20959	
/rpt_family="Alu"	
20970. .21795	
/rpt_family="L1"	
21857. .22138	
/rpt_family="Alu"	
22317. .22618	
/rpt_family="Alu"	
22672. .23037	
/rpt_family="L2"	
23749. .24050	

Mon Jun 21 09:02:12 2004

```

repeat_region      /rpt_family="MaLR"
24051..24233
repeat_region      /rpt_family="MIR"
24501..24604
repeat_region      /rpt_family="Alu"
24742..25040
repeat_region      /rpt_family="Alu"
25407..25493
repeat_region      /rpt_family="ERV1"
25582..25854
repeat_region      /rpt_family="Alu"
25866..26123
repeat_region      /rpt_family="Alu"
26679..26984
repeat_region      /rpt_family="Alu"
27085..28336
repeat_region      /rpt_family="L1"
28405..28615

Query Match      76.8%; Score 19.2; DB 9; Length 104926;
Best Local Similarity 87.5%; Pred.No. 3.3e-02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2      CCGAGGGGCGAGGGGCACATGC 25
Db      87138      CCGAGGGGCGAGGGGCACATGC 87161

RESULT 13
AC016802/c
LOCUS      AC016802.5      GI:7230035
DEFINITION      HTGS; HTGS; HTGS; DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-509
Unpublished
2 (bases 1 to 146868)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckler,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferrelita,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Harford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Kashtan,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrin,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,R., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:6980269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2879

```

Center clone name: 5\_O\_9

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 129806 bases at least Q40

Consensus quality: 138391 bases at least Q30

Consensus quality: 141610 bases at least Q20

Insert size: 79000; agarose-fp

Quality coverage: 5.8 in Q20 bases; agarose-fp

Quality coverage: 3.2 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 30 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1

1031: contig of 1031 bp in length

1032

1131: gap of 100 bp

1132

1194: contig of 1063 bp in length

2195

2194: gap of 100 bp

2295

3391: contig of 1097 bp in length

3392

3431: gap of 100 bp

3492

4032: contig of 541 bp in length

4132

4132: gap of 100 bp

4133

5982: contig of 1850 bp in length

5983

6082: gap of 100 bp

6083

7573: contig of 1491 bp in length

7574

7573: gap of 100 bp

7574

9126: contig of 1453 bp in length

9127

9226: gap of 100 bp

9227

10419: contig of 1193 bp in length

10420

10519: gap of 100 bp

10520

12629: contig of 2110 bp in length

12630

12729: gap of 100 bp

12730

14092: contig of 1363 bp in length

14093

14132: gap of 100 bp

14193

16442: contig of 2250 bp in length

16443

16542: gap of 100 bp

16543

17305: contig of 763 bp in length

17306

17405: gap of 100 bp

17406

20576: contig of 3171 bp in length

20577

20676: gap of 100 bp

20677

23407: contig of 2731 bp in length

23408

23507: gap of 100 bp

23508

26149: contig of 2642 bp in length

26150

26249: gap of 100 bp

26250

29237: contig of 2988 bp in length

29238

29337: gap of 100 bp

29338

32722: contig of 3385 bp in length

32723

32822: gap of 100 bp

32823

36212: contig of 3390 bp in length

36213

40122: contig of 3810 bp in length

40123

40222: gap of 100 bp

40223

43664: contig of 3442 bp in length

43665

43764: gap of 100 bp

43765

49178: contig of 5414 bp in length

49179

49278: gap of 100 bp

49279

53476

53575: contig of 4197 bp in length

53576

59681: gap of 100 bp

59682

59781: contig of 6106 bp in length

59782

64779: contig of 4998 bp in length

64780

64879: gap of 100 bp

64880

72046: contig of 7167 bp in length

72047

72146: gap of 100 bp

72147

81564: contig of 9418 bp in length

81565

81664: gap of 100 bp

```
* 81665 94108: contig of 12444 bp in length
* 94208: gap of 100 bp
* 108100: contig of 13892 bp in length
* 108201: gap of 100 bp
* 121190: contig of 12990 bp in length
* 121291: gap of 100 bp
* 146868: contig of 25578 bp in length.
FEATURES
  source
    1..146868
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone_lib="RP11-509"
      /clone_lib="RP11-11 Human Male BAC"
      1..1031
        /note="assembly_fragment"
      1132..2194
        /note="assembly_fragment"
      2295..3391
        /note="assembly_fragment"
      3492..4032
        /note="assembly_fragment"
      clone_end:77
        vector_side:right
      4133..5982
        /note="assembly_fragment"
      6083..7573
        /note="assembly_fragment"
      7674..9126
        /note="assembly_fragment"
      9227..10419
        /note="assembly_fragment"
      10520..12629
        /note="assembly_fragment"
      12730..14092
        /note="assembly_fragment"
      14193..16442
        /note="assembly_fragment"
      16543..17305
        /note="assembly_fragment"
      clone_end:SP6
        vector_side:left
      17406..20576
        /note="assembly_fragment"
      20677..23407
        /note="assembly_fragment"
      23508..26149
        /note="assembly_fragment"
      26250..29237
        /note="assembly_fragment"
      29338..32722
        /note="assembly_fragment"
      32823..36212
        /note="assembly_fragment"
      36313..40122
        /note="assembly_fragment"
      40223..43664
        /note="assembly_fragment"
      43765..49178
        /note="assembly_fragment"
      49279..53475
        /note="assembly_fragment"
      53576..59681
        /note="assembly_fragment"
      59782..64779
        /note="assembly_fragment"
      64880..72046
        /note="assembly_fragment"
      72147..81564
        /note="assembly_fragment"
      81665..94108
        /note="assembly_fragment"
      94209..108100
```

```
misc_feature /note="assembly_fragment"
108201..121190
misc_feature /note="assembly_fragment"
121291..146868
misc_feature /note="assembly_fragment"
ORIGIN
Query Match 76.8%; Score 19.2; DB 2; Length 146868;
Best Local Similarity 87.5%; Pred.No.3.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 CCGAGGGCCAGAGGGCAGATAGC 25
Db 133799 CCGAGGGCAGAGGGCAGATAGC 133776
RESULT 14
AC112141 166558 bp DNA linear HTG 23-FEB-2002
LOCUS Homo sapiens chromosome 4 clone RP11-291M3, WORKING DRAFT SEQUENCE,
6 unordered pieces.
ACCESSION AC112141
VERSION AC112141.4 GI:18873907
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166558)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166558)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 166558)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Feb 23, 2002 this sequence version replaced gi:18860765.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: HNH0291M03
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162870 bases at least Q40
Consensus quality: 163673 bases at least Q30
Consensus quality: 164185 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 166058; sum-of-contigs
Quality coverage: 7.48 in Q20 bases; agarose-fp
Quality coverage: 7.11 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
```

\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1513: contig of 1513 bp in length  
\* 1514: gap of unknown length  
\* 1614: contig of 2658 bp in length  
\* 4271: gap of unknown length  
\* 4372: gap of unknown length  
\* 20890: contig of 16519 bp in length  
\* 20891: gap of unknown length  
\* 64495: contig of 43505 bp in length  
\* 64496: gap of unknown length  
\* 64596: contig of 48865 bp in length  
\* 113461: gap of unknown length  
\* 113561: contig of 52998 bp in length.

FEATURES

source  
1. .166558  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/clone="RP11-291M3"  
1. .1513  
/note="assembly\_name:Contig22"  
1614. .4271  
/note="assembly\_name:Contig29"  
4372. 20890  
/note="assembly\_name:Contig30"  
20991. .64495  
/note="assembly\_name:Contig31"  
64596. .113460  
/note="assembly\_name:Contig32"  
113561. .166558  
/note="assembly\_name:Contig33"

ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 166558;  
Best Local Similarity 87.5%; Pred. No. 3.2e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGGCGCACATAGC 25  
Db 124755 CCGAGGGCCAGGCGCACATAGC 124778

RESULT 15

AC084771/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 4 clone RP11-103H17 map 4, WORKING DRAFT  
SEQUENCE, 61 unordered pieces.

AC084771  
AC084771.2 GI:11693369  
VERSION  
KEYWORDS  
HTGS PHASE1, HTGS\_DRAFT.  
SOURCE  
Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 169651)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Unpublished  
Homo sapiens chromosome 4, clone RP11-103H17

JOURNAL

REFERENCE  
2 (bases 1 to 169651)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bedalov,F., Boguslavsky,L.,  
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castie,A.,  
Choquel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeRellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,  
FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,  
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,  
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,  
Lamazaras,R., Landers,T., Lechoczky,J., Levine,R., Lieu,C., Liu,G.,  
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,  
McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,  
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,

O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,  
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,  
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Tirrell, A., Travers, M., Triglio, J., Vassiliev, H., Viel, R., Vo, A.,  
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
Zimmer, A., and Zody, M.  
Direct Submission  
Submitted (15-NOV-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 13, 2000 this sequence version replaced gi:11178112.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L10870  
Center clone name: 103\_H17  
----- Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 145723 bases at least Q40  
Consensus quality: 156808 bases at least Q30  
Consensus quality: 161002 bases at least Q20  
Insert size: 166000; agarose-fp  
Quality coverage: 2.7 in Q20 bases; agarose-fp  
Quality coverage: 3.1 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 61 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 108: contig of 108 bp in length  
\* 208: gap of 100 bp  
\* 209 848: contig of 640 bp in length  
\* 849 948: gap of 100 bp  
\* 949 1744: contig of 796 bp in length  
\* 1745 1844: gap of 100 bp  
\* 1845 2603: contig of 759 bp in length  
\* 2604 2703: gap of 100 bp  
\* 2704 3290: contig of 587 bp in length  
\* 3291 3390: gap of 100 bp  
\* 3391 3985: contig of 595 bp in length  
\* 3986 4085: gap of 100 bp  
\* 4086 4881: contig of 796 bp in length  
\* 4882 4981: gap of 100 bp  
\* 4982 5752: contig of 770 bp in length  
\* 5752 5852: contig of 100 bp  
\* 5852 6601: gap of 100 bp  
\* 6601 7729: contig of 1129 bp in length  
\* 7730 7829: gap of 100 bp  
\* 7830 8907: contig of 1078 bp in length  
\* 8908 9007: gap of 100 bp  
\* 9008 10139: contig of 1132 bp in length  
\* 10140 10239: gap of 100 bp  
\* 10240 11267: contig of 1028 bp in length  
\* 11268 11367: gap of 100 bp  
\* 11368 12384: contig of 1017 bp in length  
\* 12385 12484: gap of 100 bp  
\* 12485 13466: contig of 982 bp in length  
\* 13467 13566: gap of 100 bp

```
* 13567 14768: contig of 1202 bp in length
* 14769 14868: gap of 100 bp
* 14869 13196: contig of 18328 bp in length
* 13197 33296: gap of 100 bp
* 33297 34695: contig of 1399 bp in length
* 34696 34795: gap of 100 bp
* 34796 35982: contig of 1187 bp in length
* 35983 36082: gap of 100 bp
* 36083 37600: contig of 1518 bp in length
* 37601 37701: gap of 100 bp
* 37702 39119: contig of 1419 bp in length
* 39120 39219: gap of 100 bp
* 39220 40774: contig of 1555 bp in length
* 40775 40874: gap of 100 bp
* 40875 42141: contig of 1267 bp in length
* 42142 42241: gap of 100 bp
* 42242 43877: contig of 1636 bp in length
* 43878 43977: gap of 100 bp
* 43978 45603: contig of 1626 bp in length
* 45604 47303: gap of 100 bp
* 47304 47428: gap of 100 bp
* 47429 49104: contig of 1676 bp in length
* 49105 49204: gap of 100 bp
* 49205 51260: contig of 2056 bp in length
* 51261 51360: gap of 100 bp
* 51361 52587: contig of 1227 bp in length
* 52588 52687: gap of 100 bp
* 52688 54632: contig of 1945 bp in length
* 54633 54732: gap of 100 bp
* 54733 57117: contig of 2385 bp in length
* 57118 57217: gap of 100 bp
* 57218 59347: contig of 2129 bp in length
* 59347 61249: contig of 1803 bp in length
* 61250 61349: gap of 100 bp
* 61350 63778: contig of 2429 bp in length
* 63779 63878: gap of 100 bp
* 63879 66572: contig of 2694 bp in length
* 66573 68672: gap of 100 bp
* 68673 68757: contig of 2085 bp in length
* 68758 68857: gap of 100 bp
* 68858 72013: contig of 3156 bp in length
* 72014 72113: gap of 100 bp
* 72114 74940: contig of 2827 bp in length
* 74941 75040: gap of 100 bp
* 75041 78425: contig of 3385 bp in length
* 78426 78525: gap of 100 bp
* 78526 81262: contig of 2737 bp in length
* 81263 83433: contig of 2071 bp in length
* 83434 83533: gap of 100 bp
* 83534 87252: contig of 3719 bp in length
* 87253 87352: gap of 100 bp
* 87353 90411: contig of 3059 bp in length
* 90412 90511: gap of 100 bp
* 90512 93895: contig of 3384 bp in length
* 93896 93995: gap of 100 bp
* 93996 97729: contig of 3734 bp in length
* 97730 97829: gap of 100 bp
* 97830 100887: contig of 3058 bp in length
* 100888 100988: gap of 100 bp
* 100989 104436: contig of 3449 bp in length
* 104437 104536: gap of 100 bp
* 104537 108044: contig of 3508 bp in length
* 108045 108144: gap of 100 bp
* 108145 111767: contig of 3623 bp in length
* 111768 111867: gap of 100 bp
* 111868 115233: contig of 3366 bp in length
* 115234 115333: gap of 100 bp
* 115334 118503: contig of 3170 bp in length
* 118504 118603: gap of 100 bp
* 118604 122876: contig of 4273 bp in length

* 122877 122976: gap of 100 bp
* 122977 126997: contig of 4021 bp in length
* 126998 127097: gap of 100 bp
* 127098 131066: contig of 3969 bp in length
* 131067 131166: gap of 100 bp
* 131167 137374: contig of 6208 bp in length
* 137375 137474: gap of 100 bp
* 137475 142135: contig of 4661 bp in length
* 142136 142236: gap of 100 bp
* 142237 147072: contig of 4837 bp in length
* 147073 147172: gap of 100 bp
* 147173 153860: contig of 6688 bp in length
* 153861 153960: gap of 100 bp
* 153961 160199: contig of 6239 bp in length
* 160200 160299: gap of 100 bp
* 160300 168812: contig of 8513 bp in length
* 168813 168912: gap of 100 bp
* 168913 169651: contig of 739 bp in length.

FEATURES             Location/Qualifiers
     source            1..169651
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="4"
                        /map="4"
                        /clone="RP11-103H17"
                        /clone_lib="RPC1-11 Human Male BAC"

Query Match       76.8%; Score 19.2; DB 2; Length 169651;
Best Local Similarity 87.5%; Pred. No. 3.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CCGAGGGCCAGAGGCACATAGC 25
Db      157980 CCGAGGGCCAGAGGCACATGTC 157957

          |||||
RESULT 16
AC110113/c
LOCUS       258723 bp      DNA      linear      HTG 11-OCT-2002
DEFINITION Rattus norvegicus clone CH230-303G1, ** SEQUENCING IN PROGRESS
            *** 3 unordered pieces.
AC110113
AC110113.4 GI:23605546
VERSION     HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS    Rattus norvegicus (Norway rat)
SOURCE      Rattus norvegicus
ORGANISM    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 258723)
AUTHORS    Muzny,D,Marle., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Ayalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Deigado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
```

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, X., London, P., Longacre, S., Lopez, J., Lorensueta, L., Lozsed, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindarne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelameh, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

## Direct Submission

Unpublished

2 (bases 1 to 258723)

Worley, K.C.

Direct Submission

Submitted (10-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 258723)

Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Oct 9, 2002 this sequence version replaced gi:21738142.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

## ----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GOIN

Center clone name: CH230-303G1

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 218906 bases at least Q40

Consensus quality: 223719 bases at least Q30

Consensus quality: 226910 bases at least Q20

Estimated insert size: 223284; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 254505: contig of 254505 bp in length

\* 254506 254605: gap of unknown length

\* 254606 255840: contig of 1235 bp in length

\* 255841 255940: gap of unknown length

\* 255941 258723: contig of 2783 bp in length.

## FEATURES

## source

1. 258723

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-303G1"

## misc\_feature

1. 1320

/note="wgs end extension"

clone\_end:T7"

4253. 6051

/note="wgs end extension"

clone\_end:T7"

7039. 7884

/note="clone boundary"

clone\_end:T7"

site:MboI

end sequence:RXAKI37TJ"

26620. 29134

/note="wgs contig"

237916. 239376

/note="wgs contig"

245551. 246524

/note="clone boundary"

clone\_end:Sp6

site:MboI

end sequence:RXAKI37TV"

249513. 250744

/note="wgs end extension"

clone\_end:Sp6"

253033. 254505

/note="wgs end extension"

clone\_end:Sp6"

## ORIGIN

## Query Match

Best Local Similarity 75.2%; Score 18.8; DB 2; Length 258723;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## Qy

1 CCGGAGGGCCAGAGGGCACAT 22

|||||

Db 84855 CCAGGAGGGCCAAAGGGCACAT 84834

## RESULT 17

MMU93191/c

LOCUS

DEFINITION

Mus musculus histone deacetylase-2 (HDAC-2) gene, promoter

sequence.

U93191

U93191.1 GI:3834376

ACCESSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

273 (44), 28921-28930 (1998)

99003242

Cloning and characterization of the mouse histone deacetylase-2

gene

Zeng, Y., Tang, C.M., Yao, Y.L., Yang, W.M. and Seto, E.

1 (bases 1 to 1276)

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mus musculus

U93191

U93191.1 GI:3834376

sequence.

Mus musculus histone deacetylase-2 (HDAC-2) gene, promoter

1276 bp DNA linear ROD 04-NOV-1998

```

PUBMED 9786895
REFERENCE 2 (bases 1 to 1276)
AUTHORS Zeng,Y. and Seto,E.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1997) Molecular Oncology, H. Lee Moffitt Cancer
Center, 12902 Magnolia Drive, Tampa, FL 33612, USA
FEATURES
source
1..1276
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129/SvJ"
/db_xref="taxon:10090"
/chromosome="10"
/map="10B1"
/cell_line="ES"
1..1276
/gene="HDAC-2"
1..1276
/gene="HDAC-2"
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 1276;
Best Local Similarity 84.0%; Pred. No. 9.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGCAGCATAGC 25
Db 1128 CCGGAGGGCCGCTAGGCAGCAGC 1104

RESULT 18
AX695575 59817 bp DNA linear PAT 31-MAR-2003
LOCUS AX695575
DEFINITION Sequence 1202 from Patent WO03008583.
ACCESSION AX695575
VERSION AX695575.1 GI:29418727
KEYWORDS Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Morris,D.W. and Engelhard,E.K.
TITLE Novel compositions and methods for cancer
JOURNAL Patent: WO 03008583-A 1202 30-JAN-2003;
Sagres Discovery (US)
FEATURES
source
1..59817
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 74.4%; Score 18.6; DB 6; Length 59817;
Best Local Similarity 84.0%; Pred. No. 6.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGAGGGCAGCATAGC 25
Db 8918 CCCGAGGGCCAGCGCGCGGTAGC 8942

RESULT 19
AC100484/c
LOCUS AC100484 60042 bp DNA linear HTG 22-NOV-2001
DEFINITION Mus musculus clone RP23-141N14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC100484
VERSION AC100484.1 GI:17047850
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

# REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

```

1 (bases 1 to 60042)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-141N14
Unpublished
2 (bases 1 to 60042)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhaltier,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamaze,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,T., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1544
Center clone name: 141_N_14
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 692: contig of 692 bp in length
* 693: gap of 100 bp
* 793: contig of 735 bp in length
* 1527: gap of 100 bp
* 1627: gap of 100 bp
* 2335: contig of 708 bp in length
* 2435: gap of 100 bp
* 3154: contig of 719 bp in length
* 3254: gap of 100 bp
* 3957: contig of 713 bp in length
* 4057: gap of 100 bp
* 4767: contig of 700 bp in length
* 4867: gap of 100 bp
* 5597: contig of 730 bp in length
* 5697: gap of 100 bp
* 6418: contig of 721 bp in length
* 6518: gap of 100 bp
* 7097: contig of 579 bp in length
* 7197: gap of 100 bp
* 7940: contig of 743 bp in length
* 8040: gap of 100 bp

```

```
* * 8041 8730: contig of 750 bp in length
* * 8791 8890: gap of 100 bp
* * 8891 8933: contig of 743 bp in length
* * 9634 9733: gap of 100 bp
* * 9734 10454: contig of 721 bp in length
* * 10455 10554: gap of 100 bp
* * 10555 11273: contig of 719 bp in length
* * 11274 11373: gap of 100 bp
* * 11374 12103: contig of 730 bp in length
* * 12104 12203: gap of 100 bp
* * 12204 12941: contig of 738 bp in length
* * 12942 13041: gap of 100 bp
* * 13042 13772: contig of 731 bp in length
* * 13773 13872: gap of 100 bp
* * 13873 14588: contig of 716 bp in length
* * 14589 14688: gap of 100 bp
* * 14689 15393: contig of 705 bp in length
* * 15394 16237: contig of 744 bp in length
* * 16238 16337: gap of 100 bp
* * 16338 17071: contig of 734 bp in length
* * 17072 17171: gap of 100 bp
* * 17172 17917: contig of 746 bp in length
* * 17918 18017: gap of 100 bp
* * 18018 18727: contig of 710 bp in length
* * 18728 18827: gap of 100 bp
* * 18828 19531: contig of 704 bp in length
* * 19532 19631: gap of 100 bp
* * 19632 20349: contig of 718 bp in length
* * 20350 20449: gap of 100 bp
* * 20450 21149: contig of 700 bp in length
* * 21150 21249: gap of 100 bp
* * 21250 21978: contig of 729 bp in length
* * 21979 22078: gap of 100 bp
* * 22079 22784: contig of 706 bp in length
* * 22785 22884: gap of 100 bp
* * 22885 23636: contig of 752 bp in length
* * 23637 23736: gap of 100 bp
* * 23737 24445: contig of 709 bp in length
* * 24446 24545: gap of 100 bp
* * 24546 25292: contig of 747 bp in length
* * 25293 25392: gap of 100 bp
* * 25393 26157: contig of 765 bp in length
* * 26158 26257: gap of 100 bp
* * 26258 27004: contig of 747 bp in length
* * 27005 27104: gap of 100 bp
* * 27105 27822: contig of 718 bp in length
* * 27823 27922: gap of 100 bp
* * 27923 28548: contig of 726 bp in length
* * 28549 28748: gap of 100 bp
* * 28749 29479: contig of 731 bp in length
* * 29480 29579: gap of 100 bp
* * 29580 30325: contig of 746 bp in length
* * 30326 30425: gap of 100 bp
* * 30426 31179: contig of 754 bp in length
* * 31180 31279: gap of 100 bp
* * 31280 32022: contig of 743 bp in length
* * 32023 32122: gap of 100 bp
* * 32123 32850: contig of 728 bp in length
* * 32851 33699: contig of 749 bp in length
* * 33700 33799: gap of 100 bp
* * 33800 34506: contig of 707 bp in length
* * 34507 34607: gap of 100 bp
* * 34607 35332: contig of 726 bp in length
* * 35333 35432: gap of 100 bp
* * 35433 36158: contig of 726 bp in length
* * 36159 36258: gap of 100 bp
* * 36259 36863: contig of 605 bp in length
* * 36864 37712: contig of 749 bp in length
* * 37713 37812: gap of 100 bp
* * 37813 38555: contig of 743 bp in length
```

```
* * 38556 38655: gap of 100 bp
* * 38656 39377: contig of 722 bp in length
* * 39378 39477: gap of 100 bp
* * 39478 40181: contig of 714 bp in length
* * 40182 40291: gap of 100 bp
* * 40292 41020: contig of 729 bp in length
* * 41021 41120: gap of 100 bp
* * 41121 41810: contig of 690 bp in length
* * 41811 41910: gap of 100 bp
* * 41911 42649: contig of 739 bp in length
* * 42650 42749: gap of 100 bp
* * 42750 43498: contig of 749 bp in length
* * 43499 43598: gap of 100 bp
* * 43599 44305: contig of 707 bp in length
* * 44306 44405: gap of 100 bp
* * 44406 45156: contig of 751 bp in length
* * 45157 45256: gap of 100 bp
* * 45257 45973: contig of 717 bp in length
* * 45974 46073: gap of 100 bp
* * 46074 46782: contig of 709 bp in length
* * 46783 46882: gap of 100 bp
* * 46883 47588: contig of 706 bp in length
* * 47589 47688: gap of 100 bp
* * 47689 48406: contig of 718 bp in length
* * 48407 48506: gap of 100 bp
* * 48507 49227: contig of 721 bp in length
* * 49228 49327: gap of 100 bp
* * 49328 50057: contig of 730 bp in length
* * 50058 50157: gap of 100 bp
* * 50158 50910: contig of 753 bp in length
* * 50911 51010: gap of 100 bp
* * 51011 51756: contig of 746 bp in length
* * 51757 51856: gap of 100 bp
* * 51857 52566: contig of 710 bp in length
* * 52567 52666: gap of 100 bp
* * 52667 53402: contig of 736 bp in length
* * 53403 53502: gap of 100 bp
* * 53503 54250: contig of 748 bp in length
* * 54251 54350: gap of 100 bp
* * 54351 55050: contig of 700 bp in length
* * 55051 55150: gap of 100 bp
* * 55151 55872: contig of 722 bp in length
* * 55873 55972: gap of 100 bp
* * 55973 56691: contig of 719 bp in length
* * 56692 56791: gap of 100 bp
```

Query Match 74.4%; Score 18.6; DB 2; Length 60042;  
Best Local Similarity 84.0%; Pred. No. 6.5e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGCACATAGC 25  
Db 18332 CCCTTAGGGCTAGAGACATAGC 18308

RESULT 20  
HSIMMDL/c

LOCUS HSIMMDL 92588 bp DNA linear PRI 02-JUL-1999  
DEFINITION DNA sequence of the human immunoglobulin D segment locus.  
ACCESSION X97051.1 GI:1770449  
VERSION X97051.1  
KEYWORDS D segment; germ line; immunoglobulin heavy chain; JH segment; VH segment.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 86830 to 89967)  
Ravetch, J.V., Siebenlist, U., Korsmeyer, S., Waldmann, T. and Leder, P.  
Structure of the human immunoglobulin mu locus: characterization of  
embryonic and rearranged J and D genes  
Cell 27 (3 Pt 2), 583-591 (1981)  
JOURNAL  
MEDLINE 93358331



PUBMED  
REFERENCE 2 (bases 36339 to 36426; 45954 to 46041; 55207 to 55294; 64614 to 64700)  
AUTHORS Siebenlist,U., Ravetch,J.V., Korsmeyer,S., Waldmann,T. and Leder,P.  
TITLE Human immunoglobulin D segments encoded in tandem multigenic families  
JOURNAL Nature 294 (5842), 631-635 (1981)  
MEDLINE 82080701  
PUBMED 7312051  
REFERENCE 3 (bases 90650 to 91231)  
AUTHORS Rabbitts,T.H., Forster,A., Baer,R. and Hamlyn,P.H.  
TITLE Transcription enhancer identified near the human C mu immunoglobulin heavy chain gene is unavailable to the translocated c-myc gene in a Burkitt lymphoma  
JOURNAL Nature 306 (5945), 806-809 (1983)  
MEDLINE 84093586  
PUBMED 6419124  
REFERENCE 4 (bases 1 to 92588)  
AUTHORS Buluwela,L., Albertson,D.G., Sherrington,P., Rabbitts,P.H., Spurr,N. and Rabbitts,T.H.  
TITLE The use of chromosomal translocations to study human immunoglobulin gene organization: mapping DH segments within 35 kb of the C mu gene and identification of a new DH locus  
EMBO J. 7 (7), 2003-2010 (1988)  
JOURNAL 88328988  
MEDLINE 3138112  
PUBMED 3144456  
REFERENCE 5 (bases 12996 to 13531)  
AUTHORS Buluwela,L. and Rabbitts,T.H.  
TITLE A VH gene is located within 95 Kb of the human immunoglobulin heavy chain constant region genes  
Eur. J. Immunol. 18 (11), 1843-1845 (1988)  
JOURNAL 89078478  
MEDLINE 3144456  
PUBMED 3144456  
REFERENCE 6 (bases 38035 to 52980)  
AUTHORS Ichihara,Y., Matsuka,H. and Kurosawa,Y.  
TITLE Organization of human immunoglobulin heavy chain diversity gene loci  
EMBO J. 7 (13), 4141-4150 (1988)  
JOURNAL 89210798  
MEDLINE 3243276  
PUBMED 3243276  
REFERENCE 7 (bases 87277 to 89821)  
AUTHORS Mattila,P.S., Schugk,J., Wu,H. and Makela,O.  
TITLE Extensive allelic sequence variation in the J region of the human immunoglobulin heavy chain gene locus  
Eur. J. Immunol. 25 (9), 2578-2582 (1995)  
JOURNAL 96011870  
MEDLINE 7589129  
PUBMED 7589129  
REFERENCE 8 (bases 1 to 92588)  
AUTHORS Corbett,S.  
TITLE Direct Submission  
JOURNAL Submitted (02-APR-1996) S. Corbett, MRC, Centre for Protein Eng., Hills Road, Cambridge, CB2 2QH, UK  
COMMENT This sequence of the human immunoglobulin D segment locus was determined from three overlapping cosmids; COS23, COS21 and COS24 (described in reference [2] above). All the sequencing and analysis was carried out at the Sanger Centre (Hinxton, Cambs, CB10 1RQ, England) whose financial and technical assistance is gratefully acknowledged.  
FEATURES  
source  
1..92588  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/map="14q32.33"  
/germline  
repeat\_region <332..>429  
/note="Alu repeat: matches 99. .1 of consensus"  
repeat\_region <446..>501  
/note="Alu repeat: matches 1. .58 of consensus"  
repeat\_region <504..>600  
/note="Alu repeat: matches 211. .308 of consensus"  
repeat\_region 659..1074

repeat\_region /note="2 copies of 208 mer 82 % conserved"  
880..991  
/note="3 copies of 104 mer 97 % conserved"  
1438..1771  
/note="MUTD element fragment"  
2589..2696  
/note="L1 element fragment"  
3347..3636  
/note="Alu repeat: matches 1. .308 of consensus"  
3949..4002  
/note="27 copies of 2 mer 82 % conserved"  
3963..4002  
/note="10 copies of 4 mer 90 % conserved"  
4557..4868  
/note="3 copies of 104 mer 82 % conserved"  
4939..5227  
/note="Alu repeat: matches 1. .308 of consensus"  
5830..6197  
/note="2 copies of 184 mer 88 % conserved"  
5920..5983  
/note="Alu repeat: matches 1. .180 of consensus"  
9385..9453  
/note="THE1B element fragment"  
9389..9468  
/note="WSTA element fragment"  
9592..10057  
/note="3 copies of 22 mer 89 % conserved"  
13036..91231  
/gene="IGH@"  
join(13036..13084,13168..>13481)  
/gene="IGH@"  
/citation=[5]  
/codon\_start=1  
/product="immunoglobulin heavy chain"  
/protein\_id="CAA65763.1"  
/db\_xref="GI:1813977"  
/db\_xref="RENTREMBL:CAA65763"  
/translation="MSVSFLIFLPVLGPMGLVSOVLQSGFGLVKPQSTLSLTCAI  
SGDSVSSNAWNLRQSPSRGLEWLGRTYRSKWYNDYAVSVKSRITINPTDSKNQF  
SGLNSVTPEDTAVYICAR"  
join(13036..13084,13168..13178)  
/gene="IGH@"  
/citation=[5]  
13179..>13481  
/gene="IGH@"  
/product="immunoglobulin heavy chain"  
/citation=[5]  
13085..13167  
/gene="IGH@"  
/note="VH6-1 leader intron"  
/citation=[5]  
13179..13483  
/gene="IGH@"  
/product="immunoglobulin heavy chain"  
/standard\_name="VH6-1"  
/citation=[5]  
13856..14008  
/note="L1 element fragment"  
16622..16883  
/note="WSTA element fragment"  
16636..16704  
/note="THE1B element fragment"  
16757..16977  
/note="THE1B element fragment"  
18114..18345  
/note="L1 element fragment"  
18445..18743  
/note="L1 element fragment"  
18813..>18987  
/note="Alu repeat: matches 308. .133 of consensus"  
19004..>19115  
/note="Alu repeat: matches 113. .1 of consensus"  
19371..19470

```

repeat_region /note="Li element fragment"
20192..20275
/note="2 copies of 42 mer 94 % conserved"
repeat_region 20282..20365
/note="2 copies of 42 mer 94 % conserved"
repeat_region 21647..21739
/note="2 copies of 76 mer 98 % conserved"
repeat_region 21726..21939
/note="3 copies of 78 mer 98 % conserved"
repeat_region 22161..22382
/note="Li element fragment"
22752..23135
/note="8 copies of 48 mer 89 % conserved"
repeat_region 25123..25301
/note="MT1D element fragment"
25695..25760
/note="3 copies of 22 mer 83 % conserved"
repeat_region 25697..25760
/note="16 copies of 4 mer 83 % conserved"
repeat_region 25700..25763
/note="4 copies of 16 mer 83 % conserved"
repeat_region 28283..28318
/note="12 copies of 3 mer 83 % conserved"
repeat_region 29990..30265
/note="2 copies of 138 mer 81 % conserved"
repeat_region 30572..30617
/note="MT1A element fragment"
30683..30712
/note="MT1A element fragment"
30801..30872
/note="MSTC element fragment"
31700..31825
/note="3 copies of 42 mer 99 % conserved"
D_segment 33714..33730
/gene="IGH"
/product="immunoglobulin heavy chain"
/note="DI-1"
/citation=[8]

```

```

Query Match 74.4%; Score 18.6; DB 9; Length 92588;
Best Local Similarity 84.0%; Pred. No. 6.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGCCGAGGGGCACATAGC 25
Db 69447 CACAGAGGCCGAGGGGCACAGC 69423

```

```

RESULT 21
AL929178 AL929178 100057 bp DNA linear ROD 18-JAN-2003
LOCUS Mouse DNA sequence from clone RP23-77M17 on chromosome 2, complete
DEFINITION sequence.
ACCESSION AL929178
VERSION AL929178.7 GI:27803212
KEYWORDS HTG
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 100057)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Glithero,R.
JOURNAL Direct Submission
COMMENT Submitted (18-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jan 18, 2003 this sequence version replaced gi:27764108.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
-----
Center: Wellcome Trust Sanger Institute

```

```

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information
on the WormPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep RP23-77M17 is
from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source           1..100057
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-77M17"
                     /clone_lib="RPCI-23"
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 100057;
Best Local Similarity 84.0%; Pred. No. 6.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGCCGAGGGGCACATAGC 25
Db 22573 CCCTTAGGCTAGAGAGACATAGC 22597

RESULT 22
AL603889/c AL603889 145052 bp DNA linear ROD 05-APR-2002
LOCUS Mouse DNA sequence from clone RP23-407M20 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL603889
VERSION AL603889.5 GI:20068524
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Van Hellmond,Z.
JOURNAL Direct Submission
COMMENT Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:19031503.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate

```



database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-18012 is  
 from the RPCI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACE3.6.

## FEATURES

source  
 1..151555  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="4"  
 /clone="RP23-18012"  
 /clone\_lib="RPCI-23"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 151555;  
 Best Local Similarity 84.0%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGCCGACAGGCGCACATAGC 25

Db 140146 CCCTTAGGGCTAGAGCACATAGC 140170

## RESULT 25

AB019441/C

LOCUS

DEFINITION Homo sapiens DNA for immunoglobulin heavy-chain variable region,  
 complete sequence, 5 of 5.

ACCESSION AB019441

VERSION

KEYWORDS immunoglobulin heavy chain variable region; VH.

SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H.,  
 Miyata, T. and Honjo, T.  
 The complete nucleotide sequence of the human immunoglobulin heavy  
 chain variable region locus

J. Exp. Med. 188 (11), 2151-2162 (1998)

99059838

PUBMED

9841928

REFERENCE 2 (bases 1 to 157090)

AUTHORS

Matsuda, F.

Direct Submission

Submitted (05-NOV-1998) Fumihiko Matsuda, Kyoto University Graduate  
 School of Medicine, Department of Medical Chemistry, Yoshida  
 Konoe-cho, Sakyo-ku, Kyoto 606, Japan  
 (E-mail: fmatsuda@viru1.virus.kyoto-u.ac.jp, Tel: 81-75-753-4371,  
 Fax: 81-75-753-4388)

## FEATURES

source

1..157090  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /map="14q32.33"  
 /map="10920..10927"

/note="octameric regulatory sequence"

10936..10942

11002..83932

/gene="VH"

/pseudo

join(11002..11047,11130..>1433)

/gene="VH"

/note="4-4"

/pseudo

/codon\_start=1

/product="immunoglobulin heavy chain variable region"

/protein\_id="BAA75058.1"

exon <11002..11047  
 /gene="VH"  
 /number=1  
 /pseudo  
 intron 11048..11129  
 /gene="VH"  
 exon 11130..11433  
 /gene="VH"  
 /number=2  
 /pseudo  
 misc\_feature 11434..11440  
 /gene="VH"  
 /note="7mer RSS: recombination signal sequence cacagtg"  
 /pseudo 11464..11472  
 /gene="VH"  
 /note="9mer RSS: recombination signal sequence acataaaccc"  
 /pseudo 11751..11757  
 /gene="VH"  
 /note="heptameric regulatory sequence"  
 /pseudo 11760..11767  
 /gene="VH"  
 /note="octameric regulatory sequence"  
 /pseudo 11776..11782  
 /gene="VH"  
 /note="1-3"  
 /codon\_start=1  
 /product="immunoglobulin heavy chain variable region"  
 /protein\_id="BAA75059.1"  
 /db\_xref="GI:4512313"  
 <17860..17905  
 /gene="VH"  
 /number=1  
 /pseudo 17906..17990  
 /gene="VH"  
 17991..18297  
 /gene="VH"  
 /number=2  
 /pseudo 18298..18304  
 /gene="VH"  
 /note="7mer RSS: recombination signal sequence cacagtg"  
 /pseudo 18328..18336  
 /gene="VH"  
 /note="9mer RSS: recombination signal sequence tcaagaacc"  
 /pseudo <21563..>21864  
 /gene="VH"  
 /note="3-02.1P"  
 /pseudo  
 /codon\_start=1  
 /product="immunoglobulin heavy chain variable region"  
 36321..36327  
 /gene="VH"  
 /note="heptameric regulatory sequence"  
 /pseudo 36330..36337  
 /gene="VH"  
 /note="octameric regulatory sequence"  
 /pseudo 36346..36352  
 /gene="VH"  
 /note="1-3"  
 /codon\_start=1  
 /product="immunoglobulin heavy chain variable region"  
 /protein\_id="BAA75058.1"  
 >36872)  
 /gene="VH"

sequence.  
AL645923  
VERSION AL645923.14 GI:19572045  
HTG.  
Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 157361)  
AUTHORS Smith, M.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
COMMENT On Mar 21, 2002 this sequence version replaced gi.19335835. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, ENBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-75H1 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACE3.6  
-----Genome Center  
Center: UK Medical Research Council  
Center code: UK-MRC  
Web site: <http://mrseq.har.mrc.ac.uk>  
Contact: mouse@har.mrc.ac.uk  
-----  
FEATURES  
source  
1..157361  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="13"  
/clone="RP23-75H1"  
/clone\_lib="RPCI-23"  
ORIGIN  
Query Match 74.4%; Score 18.6; DB 10; Length 157361;  
Best Local Similarity 84.0%; Pred. No. 5.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 CCGGAGGGCCAGAGGGCACATAGC 25  
Db 116583 CCTTAGGGCTAGAGGACATAGC 116607  
RESULT 27  
AC092729/c  
LOCUS AC092729 165116 bp DNA linear MAM 28-OCT-2003  
DEFINITION Canis familiaris clone RP81-60B6, complete sequence.  
ACCESSION AC092729  
VERSION AC092729.3 GI:37999253  
KEYWORDS HTG.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

/note="1-2"  
/pseudo  
/codon\_start=1  
/product="immunogloblin heavy chain variable region"  
/protein\_id="BAA75060.1"  
/db\_xref="GI:4512314"  
<36435..36480  
/gene="VH"  
/number=1  
/pseudo  
36481..36565  
/gene="VH"  
36566..36872  
/gene="VH"  
/number=2  
/pseudo  
36873..36879  
/gene="VH"  
/note="7mer RSS: recombination signal sequence cacagtg"  
/pseudo  
36903..36911  
/gene="VH"  
/note="9mer RSS: recombination signal sequence tcagaacc"  
/pseudo  
<78294..>78475  
/gene="VH"  
/note="4-01.1p"  
/pseudo  
/codon\_start=1  
/product="immunogloblin heavy chain variable region"  
83384..83391  
/gene="VH"  
/note="octameric regulatory sequence"  
/pseudo  
83400..83406  
/gene="VH"  
Join(83485..83533,83617..>83932)  
/gene="VH"  
/note="6-1"  
/pseudo  
/codon\_start=1  
/product="immunogloblin heavy chain variable region"  
/protein\_id="BAA75061.1"  
/db\_xref="GI:4512315"  
<83485..83533  
/gene="VH"  
/number=1  
/pseudo  
83534..83616  
/gene="VH"  
83617..83932  
/gene="VH"  
/number=2  
/pseudo  
83933..83939  
/note="7mer RSS: recombination signal sequence cacagtg"  
83963..83971  
/note="9mer RSS: recombination signal sequence acacaacc"  
ORIGIN  
Query Match 74.4%; Score 18.6; DB 9; Length 157090;  
Best Local Similarity 84.0%; Pred. No. 5.9e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 CCGGAGGGCCAGAGGGCACATAGC 25  
Db 139905 CACAGAGGGCCAGAGGGCACATAGC 139881  
RESULT 26  
AL645923  
LOCUS AL645923 157361 bp DNA linear ROD 29-JUN-2002  
DEFINITION Mouse DNA sequence from clone RP23-75H1 on chromosome 13, complete



LOCUS CNS01D72 169802 bp DNA linear PRI 19-NOV-2001  
 DEFINITION Human chromosome 14 DNA sequence BAC R-417P24 of library RPCI-11  
 from chromosome 14 of Homo sapiens (Human), complete sequence.  
 ACCESSION AL122127  
 VERSION AL122127.6 GI:17026193  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (Bases 1 to 169802)  
 AUTHORS Heilig, R., Petit, J. L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,  
 Brothier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,  
 Levy, M., Eckenberg, R., Bruls, T., DeBerardinis, V., Cruaud, C.,  
 Gysay, G., Saurin, W. and Weissenbach, J.  
 TITLE Sequencing of the human chromosome 14  
 JOURNAL Unpublished  
 REFERENCE 2 (Bases 1 to 169802)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-NOV-2001) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT On Nov 20, 2001 this sequence version replaced gi:14715169.  
 ----- Genom Center  
 Center: Genoscope / Centre National de Sequencage  
 Center code: GS  
 Web site: <http://www.genoscope.cns.fr/>  
 Contact: Seqref@genoscope.cns.fr  
 -----  
 The following BAC sequence is oriented from the T7 to the SP6 end.  
 ----- Summary Statistics  
 Assembly program: Phrap; version 2.0  
 Quality coverage: 7.56x in Q20 bases; sum-of-contigs  
 -----  
 Overall quality chart :  
 Range : bases  
 0 :  
 1 - 9 :  
 10 - 19 : 11  
 20 - 29 : 159  
 30 - 39 : 4325  
 40 - 49 : 11964  
 50 - 59 : 13264  
 60 - 69 : 28113  
 70 - 79 : 55713  
 80 - 89 : 56253  
 90 - 99 :  
 -----  
 Percentage of bases with a quality value >= 40 : 99 %.  
 Location/Qualifiers  
 1. 169802  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone\_lib="RPCI-11"  
 /clone="R-417P24"  
 11427..11641  
 /note="matching EMBL:G33053  
 RHdb:RH67749  
 dbSTS:STS47676  
 Identified using the e-PCR software (G. Schuler)"  
 11439..11536  
 /note="matching EMBL:G14654  
 RHdb:RH7974  
 dbSTS:STS21065  
 Identified using the e-PCR software (G. Schuler)"  
 39191..39405  
 /note="matching EMBL:G33053  
 RHdb:RH67749  
 dbSTS:STS47676

Identified using the e-PCR software (G. Schuler)"  
 39203..39300  
 /note="matching EMBL:G14654  
 RHdb:RH7974  
 dbSTS:STS21065  
 Identified using the e-PCR software (G. Schuler)"  
 39281..39447  
 /note="matching EMBL:X99549  
 RHdb:RH71306  
 dbSTS:STS51071  
 Identified using the e-PCR software (G. Schuler)"  
 131895..132128  
 /note="matching EMBL:X00253  
 RHdb:RH18069  
 dbSTS:STS3713  
 Identified using the e-PCR software (G. Schuler)"  
 147503..147627  
 /note="matching EMBL:R05773  
 RHdb:RH53956  
 dbSTS:STS23199  
 Identified using the e-PCR software (G. Schuler)"  
 ORIGIN  
 Query Match 74.4%; Score 18.6; DB 9; Length 169802;  
 Best Local Similarity 84.0%; Pred.No. 5.8e+02;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CCGGAGGGCCAGAGGGCCATAGC 25  
 Db 153330 CACAGAGGGCCAGAGGGCCAGACAGC 153354  
 RESULT 30  
 BX546505/c  
 LOCUS  
 DEFINITION Mouse DNA sequence from clone Rp23-220E8 on chromosome X, complete  
 sequence.  
 ACCESSION BX546505  
 VERSION BX546505.8 GI:37606195  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (Bases 1 to 183564)  
 AUTHORS Chapman, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Oct 9, 2003 this sequence version replaced gi:35209624.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.  
 ----- Genom Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: humquery@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the





OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 143.651 Seconds  
(without alignments)  
739.327 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccggaggccagagggcacatagc 25

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 337863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N Geneseq 29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001as:\*

5: Geneseq2002as:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2003cs:\*

9: Geneseq2004s:\*

10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	100.0	1584	4	AaH51956 Mycobacte
C 2	25	100.0	110000	4	AaI99682 Mycobacte
C 3	25	100.0	110000	4	AaI99683 Mycobacte
C 4	18.6	74.4	59817	8	Ada02884 Human DUS
C 5	18.6	74.4	59817	9	Ada72422 Human DUS
C 6	18.4	73.6	14646	6	AbK64796 Human ben
C 7	18.2	72.8	1409	4	AbI05531 Drosophil
C 8	18.2	72.8	2696	4	AbI05332 Drosophil
C 9	18.2	72.8	2722	4	AbI05344 Drosophil
C 10	18.2	72.8	3409	4	AbI05530 Drosophil
C 11	18.2	72.8	90442	8	Ada03077 Mouse mCG
C 12	18.2	72.8	90442	8	Ada66361 Mouse mCG
C 13	18.2	72.8	90442	9	Ada72815 Mouse mCG
C 14	17.6	70.4	492	6	AaS98106 Human DNA
C 15	17.6	70.4	637	4	AaS42975 DNA encod
C 16	17.6	70.4	993	6	AaS98057 Human DNA
C 17	17.6	70.4	1152	9	Ada86886 Human GPC
C 18	17.6	70.4	3311	9	Ada28123 Human NTR
C 19	17.2	68.8	373	4	AaK88002 Human dig
C 20	17.2	68.8	2000	7	Ada73174 Rice gene
C 21	17.2	68.8	2266	4	AaK89548 Human dig
C 22	17.2	68.8	10733	6	AbI99772 Mouse isc
C 23	17.2	68.8	16038	4	AaK89507 Human dig

24	17.2	68.8	16595	2	AAX23521	Aax23521 Human kid
25	17.2	68.8	23070	8	ADA02507	Ada02507 Mouse Wnt
26	17.2	68.8	23070	9	ADB72245	Adb72245 Mouse Wnt
C 27	17	68.0	376	3	AAF18390	Aaf18390 Lung canc
C 28	17	68.0	406	8	ACH19776	Ach19776 Human adu
C 29	17	68.0	437	3	AAF18383	Aaf18383 Lung canc
C 30	17	68.0	547	6	ABL88406	AbI88406 Pain regu
C 31	17	68.0	552	4	AAI19014	Aai19014 Probe #89
C 32	17	68.0	552	4	ABA64017	AbA64017 Human foe
C 33	17	68.0	552	4	ABA31186	AbA31186 Probe #36
C 34	17	68.0	552	4	AAK38226	Aak38226 Human bon
C 35	17	68.0	552	4	ABS12265	AbS12265 Human gen
C 36	17	68.0	681	9	ADD49124	Add49124 Human NOV
C 37	17	68.0	941	1	AA71010	Aa71010 Sequence
C 38	17	68.0	941	2	AAQ02893	Aaq02893 cDNA inse
C 39	17	68.0	948	1	AA71009	Aa71009 Sequence
C 40	17	68.0	948	2	AAQ02892	Aaq02892 cDNA inse
C 41	17	68.0	997	4	AAI18617	Aai18617 Probe #85
C 42	17	68.0	997	4	ABA63615	AbA63615 Human foe
C 43	17	68.0	997	4	AAI43731	Aai43731 Probe #12
C 44	17	68.0	997	4	ABA30812	AbA30812 Probe #32
C 45	17	68.0	997	4	AAK37857	Aak37857 Human bon
C 46	17	68.0	997	4	AAK12141	Aak12141 Human bra
C 47	17	68.0	997	4	ABS37486	AbS37486 Human liv
C 48	17	68.0	997	6	ABS11852	AbS11852 Human gen
C 49	17	68.0	1159	2	AAQ04080	Aaq04080 Vector pA
C 50	17	68.0	1161	1	AAN60379	Aan60379 Plasmid p
C 51	17	68.0	1161	1	AAN60571	Aan60571 Sequence
C 52	17	68.0	1161	9	ADA49126	Add49126 Human NOV
C 53	17	68.0	1836	7	ACA39969	AcA39969 Prokaryot
C 54	17	68.0	2139	9	ADC32008	AdC32008 Human nov
C 55	17	68.0	2248	3	AAF18011	Aaf18011 Lung canc
C 56	17	68.0	2717	1	AAN60380	Aan60380 Plasmid p
C 57	17	68.0	2719	1	AAN60572	Aan60572 Genomic s
C 58	17	68.0	2720	2	AAQ04081	Aaq04081 Vector pA
C 59	17	68.0	2759	1	AAN60570	Aan60570 Genomic s
C 60	17	68.0	2761	2	AAQ04082	Aaq04082 Vector pM
C 61	17	68.0	4059	9	ADE55268	AdE55268 Rat gene
C 62	17	68.0	4069	9	ADA45166	Add45166 Rat gene
C 63	17	68.0	4069	9	AD55272	AdE55272 Rat gene
C 64	17	68.0	4069	9	AD55276	AdE55276 Rat gene
C 65	17	68.0	4069	9	AD55280	AdE55280 Rat gene
C 66	17	68.0	4585	2	AAQ04077	Aaq04077 Human 32K
C 67	17	68.0	4732	6	ABU66769	AbU66769 Lung canc
C 68	17	68.0	4732	6	ABK64546	AbK64546 Human ben
C 69	17	68.0	4778	6	ABL66721	AbL66721 Lung canc
C 70	17	68.0	5233	3	AAC63530	Aac63530 Slit-2 pr
C 71	17	68.0	5700	5	AAS66584	Aas66584 DNA encod
C 72	17	68.0	7852	5	ABA14491	AbA14491 Human ner
C 73	17	68.0	9685	4	AAK5074	Aak5074 Human imm
C 74	17	68.0	9693	4	AAK95075	Aak95075 Human imm
C 75	17	68.0	14928	2	AAQ11707	Aaq11707 Immunoglo
C 76	17	68.0	21732	4	ABK42527	AbK42527 Genomic s
C 77	17	68.0	21732	4	AAK81131	Aak81131 Human imm
C 78	17	68.0	21732	5	ABA14492	AbA14492 Human ner
C 79	17	68.0	21732	8	ADB60683	AdB60683 Connectiv
C 80	17	68.0	37160	6	ABK84372	AbK84372 Human CDN
C 81	17	68.0	174424	6	ABL68122	AbL68122 Ovary can
C 82	17	68.0	186510	9	ADE24797	AdE24797 Human end
C 83	17	68.0	201143	6	ABK83568	AbK83568 Human DNA
C 84	16.8	67.2	462	4	AAS30749	Aas30749 Human CDN
C 85	16.8	67.2	484	8	ACH44098	Ach44098 Human foe
C 86	16.8	67.2	676	4	AAI10066	Aai10066 Human bre
C 87	16.8	67.2	812	4	AAI19646	Aai19646 Human bre
C 88	16.8	67.2	987	5	AAF28251	Aaf28251 Human TAN
C 89	16.8	67.2	1317	6	AAS62393	Aas62393 cDNA sequ
C 90	16.8	67.2	1868	4	AAK51955	Aak51955 Human pol
C 91	16.8	67.2	1926	4	AAH14931	Aah14931 Human CDN
C 92	16.8	67.2	1926	9	ADD19037	Add19037 Human dis
C 93	16.8	67.2	1976	4	ABA05041	AbA05041 Human sec
C 94	16.8	67.2	1976	4	AAK52939	Aak52939 Human pol
C 95	16.8	67.2	1979	3	AAA95784	Aaa95784 Human imm
C 96	16.8	67.2	1979	8	ACH04022	Ach04022 Human CDN

97	16.8	67.2	1992	5	AAE28250	Aar28250 Human TAN	170	16.2	64.8	560	6	ABQ47012	Abq47012 Oligonucle
98	16.8	67.2	2017	4	AAI26420	Aai26420 Human bre	c 171	16.2	64.8	560	6	ABQ47013	Abq47013 Oligonucle
99	16.8	67.2	2017	4	AAH26622	Aah26622 Human cer	172	16.2	64.8	567	7	ACD93788	Acd93788 Human col
100	16.8	67.2	2017	4	AAH26637	Aah26637 Human cer	c 173	16.2	64.8	709	7	ABX05183	Abx05183 Human nov
101	16.8	67.2	2038	6	AB235495	Abx235495 Human gen	c 174	16.2	64.8	788	5	ABA21294	Abx21294 Human rer
102	16.8	67.2	2038	7	ABX622913	Abx622913 Human act	175	16.2	64.8	931	3	AAE69549	Aae69549 Human rer
c 103	16.8	67.2	2059	2	ACH00141	Ach00141 Hypoxia-r	176	16.2	64.8	1262	3	AAZ98050	Aaz98050 Human sec
c 104	16.8	67.2	2059	8	ACH00174	Ach00174 Rat parti	177	16.2	64.8	1262	4	AAI11663	Aai11663 Human sec
c 105	16.8	67.2	2059	9	ADC69804	Adc69804 Rat Rtp95	178	16.2	64.8	1262	6	ABK69759	Abk69759 Human sec
c 106	16.8	67.2	2059	9	ADG62297	Adg62297 Hypoxia-r	179	16.2	64.8	1262	7	ADA55902	Ada55902 Gene enco
c 107	16.8	67.2	2624	9	ADG62642	Adg62642 Rat gene	180	16.2	64.8	1262	7	ADA39708	Ada39708 Human sec
c 108	16.8	67.2	2624	9	ADG62646	Adg62646 Rat gene	181	16.2	64.8	1282	9	ADC73437	Adc73437 Human sec
c 109	16.8	67.2	3156	6	AB211388	Abx211388 Human pol	182	16.2	64.8	1357	2	AAI18575	Aai18575 Human tum
c 110	16.8	67.2	3458	7	ADA53359	Ada53359 Human cod	183	16.2	64.8	1357	2	AAI18584	Aai18584 Human tum
c 111	16.8	67.2	3671	5	ADA02591	Ada02591 Human cyt	184	16.2	64.8	1500	7	ACF64372	Acf64372 Human TNF
c 112	16.8	67.2	13240	9	ADC86880	Adc86880 Human GPC	c 185	16.2	64.8	1560	5	ABA21304	Abx21304 Human ner
c 113	16.8	67.2	13240	9	ADC86880	Adc86880 Human GPC	c 186	16.2	64.8	1560	5	ABA21304	Abx21304 Human ner
c 114	16.8	67.2	15254	9	ADC86802	Adc86802 Human GPC	187	16.2	64.8	1862	8	ADB32793	Adb32793 Human tum
c 115	16.6	66.4	364	3	AAC99148	Aac99148 Human pan	188	16.2	64.8	2475	4	AAK34903	Aak34903 Human ful
c 116	16.6	66.4	533	2	AAH14568	Aah14568 CD11b gen	189	16.2	64.8	2654	1	AAH70597	Aah70597 Sequence
c 117	16.6	66.4	533	3	AAH35031	Aah35031 Human ade	190	16.2	64.8	2714	7	ADAI13376	Adai13376 Human int
c 118	16.6	66.4	533	3	AAH35031	Aah35031 Human low	191	16.2	64.8	2868	5	AAH74008	Aah74008 DNA enco
c 119	16.6	66.4	533	7	ABZ96847	Abz96847 Human nuc	192	16.2	64.8	2945	3	AAE68095	Aae68095 Human sec
c 120	16.6	66.4	621	5	AAH84646	Aah84646 DNA enco	193	16.2	64.8	2945	4	AAH33162	Aah33162 Human col
c 121	16.6	66.4	1485	3	AAC49413	Aac49413 Arabidops	194	16.2	64.8	2945	7	ABZ67091	Abz67091 Human sec
c 122	16.6	66.4	1496	3	AAH33585	Aah33585 Arabidops	195	16.2	64.8	2945	7	ABZ73482	Abz73482 Secrete
c 123	16.6	66.4	1774	3	AAH35033	Aah35033 Human ade	c 196	16.2	64.8	3867	2	AAV28111	Aav28111 Rat 5-oxo
c 124	16.6	66.4	1774	3	AAH35033	Aah35033 Human low	c 197	16.2	64.8	4003	9	ADB58822	Adb58822 Toxicity-
c 125	16.6	66.4	1774	7	ABZ96849	Abz96849 Human nuc	c 198	16.2	64.8	4016	2	AAV28112	Aav28112 Rat 5-oxo
c 126	16.6	66.4	1774	7	ABZ96849	Abz96849 Human rep	c 199	16.2	64.8	4171	4	AAH06355	Aah06355 Human rep
c 127	16.6	66.4	1774	4	ABL02444	Abi02444 Drosophil	c 200	16.2	64.8	4669	6	ABT10922	Abt10922 Human bre
c 128	16.6	66.4	3679	4	AAK52859	Aak52859 Human pol	c 201	16.2	64.8	5168	5	AAH81550	Aah81550 DNA enco
c 129	16.6	66.4	3794	5	AAH93067	Aah93067 Human enco	c 202	16.2	64.8	5610	7	ADH83845	Adh83845 Human COL
c 130	16.6	66.4	3999	4	ABU17929	Abu17929 Drosophil	c 203	16.2	64.8	5714	6	ABQ60955	Abq60955 Bullicus p
c 131	16.6	66.4	4237	5	AAH86660	Aah86660 DNA enco	c 204	16.2	64.8	6695	4	AAK70340	Aak70340 Human imm
c 132	16.6	66.4	6910	4	ABU17928	Abu17928 Drosophil	c 205	16.2	64.8	6695	4	AAK82282	Aak82282 Human imm
c 133	16.6	66.4	6981	6	ABH86588	Abh86588 Rat NG2 p	c 206	16.2	64.8	6911	6	AAH45858	Aah45858 Human tum
c 134	16.6	66.4	7047	3	AAH35034	Aah35034 Human ade	c 207	16.2	64.8	6911	6	AAH45858	Aah45858 Human tum
c 135	16.6	66.4	7047	3	AAH35034	Aah35034 Human low	c 208	16.2	64.8	7112	4	AAH86085	Aah86085 Lymphotox
c 136	16.6	66.4	7047	3	AAH35034	Aah35034 Human nuc	c 209	16.2	64.8	7112	4	AAH86085	Aah86085 Lymphotox
c 137	16.6	66.4	10884	2	AAH35034	Aah35034 Human rep	c 210	16.2	64.8	7112	5	AAH86085	Aah86085 Lymphotox
c 138	16.6	66.4	11987	4	AAH35034	Aah35034 Human tes	c 211	16.2	64.8	7112	7	AAH57450	Aah57450 Human tum
c 139	16.6	66.4	11987	4	AAH35034	Aah35034 Human tes	c 212	16.2	64.8	14063	3	AAH57450	Aah57450 Human tum
c 140	16.6	66.4	12575	4	AAH35034	Aah35034 Human tes	c 213	16.2	64.8	16310	3	AAH34964	Aah34964 Human ade
c 141	16.6	66.4	12575	4	AAH35034	Aah35034 Human tes	c 214	16.2	64.8	16310	3	AAH34964	Aah34964 Human ade
c 142	16.6	66.4	17514	9	AAH85026	Aah85026 Human imm	c 215	16.2	64.8	17634	3	AAH34964	Aah34964 Human ade
c 143	16.6	66.4	28564	9	AAH85026	Aah85026 Human imm	c 216	16.2	64.8	17634	3	AAH34964	Aah34964 Human ade
c 144	16.6	66.4	43793	8	AAH85026	Aah85026 Human imm	c 217	16.2	64.8	17634	3	AAH34964	Aah34964 Human ade
c 145	16.6	66.4	43793	9	AAH85026	Aah85026 Human imm	c 218	16.2	64.8	32217	7	AAH85026	Aah85026 Human imm
c 146	16.6	66.4	43794	9	AAH85026	Aah85026 Human imm	c 219	16.2	64.8	81800	6	AAH85026	Aah85026 Human imm
c 147	16.6	66.4	90442	9	AAH85026	Aah85026 Human imm	c 220	16.2	64.8	146547	7	AAH85026	Aah85026 Human imm
c 148	16.6	66.4	340449	9	AAH85026	Aah85026 Human imm	c 221	16.2	64.8	201	7	ABX55338	Abx55338 Bovine BS
c 149	16.4	65.6	261	6	ABN76364	Abn76364 Human ORF	c 222	16.2	64.8	218	4	AAI82706	Aai82706 Human pol
c 150	16.4	65.6	276	6	ABN76364	Abn76364 Human ORF	c 223	16.2	64.8	218	4	AAI82706	Aai82706 Human pol
c 151	16.4	65.6	480	9	ABN23078	Abn23078 Human ORF	c 224	16.2	64.8	294	4	AAI22459	Aai22459 Probe #12
c 152	16.4	65.6	480	9	ABN23078	Abn23078 Human ORF	c 225	16.2	64.8	294	4	ABX67534	Abx67534 Human fee
c 153	16.4	65.6	480	9	ABN23078	Abn23078 Human ORF	c 226	16.2	64.8	294	4	AAI47751	Aai47751 Probe #16
c 154	16.4	65.6	575	7	ABZ59001	Abz59001 Human gen	c 227	16.2	64.8	294	4	ABX49633	Abx49633 Human bre
c 155	16.4	65.6	577	7	ABZ59001	Abz59001 Human gen	c 228	16.2	64.8	294	4	ABX49633	Abx49633 Human bre
c 156	16.4	65.6	914	7	ABZ59001	Abz59001 Human gen	c 229	16.2	64.8	294	4	AAK41706	Aak41706 Human bon
c 157	16.4	65.6	1494	9	ABN38895	Abn38895 Mouse pro	c 230	16.2	64.8	294	4	AAK41706	Aak41706 Human bon
c 158	16.4	65.6	1950	6	ABQ73820	Abq73820 Human col	c 231	16.2	64.8	294	5	ABSA11295	Abx11295 Human liv
c 159	16.4	65.6	2240	6	ABN84376	Abn84376 Mouse var	c 232	16.2	64.8	294	4	AAI08146	Aai08146 Probe #81
c 160	16.4	65.6	2240	6	ABN84376	Abn84376 Mouse var	c 233	16.2	64.8	294	6	ABSA11295	Abx11295 Human liv
c 161	16.4	65.6	2464	7	ABZ58998	Abz58998 Human onc	c 234	16.2	64.8	298	2	ABSA11295	Abx11295 Human liv
c 162	16.4	65.6	3761	7	ABZ58998	Abz58998 Human onc	c 235	16.2	64.8	298	2	ABSA11295	Abx11295 Human liv
c 163	16.4	65.6	3761	7	ABZ58998	Abz58998 Human onc	c 236	16.2	64.8	298	2	ABSA11295	Abx11295 Human liv
c 164	16.4	65.6	4077	9	ABZ58998	Abz58998 Human onc	c 237	16.2	64.8	346	4	AAK88203	Aak88203 Human dig
c 165	16.4	65.6	10735	7	ABZ58998	Abz58998 Human onc	c 238	16.2	64.8	346	5	AAK88203	Aak88203 Human dig
c 166	16.2	64.8	291	2	AAH21652	Aah21652 Human bre	c 239	16.2	64.8	364	5	ABV51053	Abv51053 Human pro
c 167	16.2	64.8	319	6	ABN17696	Abn17696 Human gen	c 240	16.2	64.8	404	5	ABV51053	Abv51053 Human pro
c 168	16.2	64.8	455	5	AAH84998	Aah84998 DNA enco	c 241	16.2	64.8	431	7	AAH84998	Aah84998 DNA enco
c 169	16.2	64.8	473	6	ABN73488	Abn73488 Bovine em	c 242	16.2	64.8	438	6	AAH84998	Aah84998 DNA enco

## ALIGNMENTS

RESULT 1	AAH51956/c	AAH51956 standard; DNA; 1584 BP.
ID	AAH51956	
XX	AC	
XX	AC	
XX	AC	
DT	04-SEP-2001	(first entry)
XX		
XX		
DE	Mycobacterium tuberculosis	potential drug target gene SEQ ID 10.
XX		

WP AA199682\_17 1700001 1810000  
 WP AA199682\_18 1800001 1910000  
 WP AA199682\_19 1900001 2010000  
 WP AA199682\_20 2000001 2110000  
 WP AA199682\_21 2100001 2210000  
 WP AA199682\_22 2200001 2310000  
 WP AA199682\_23 2300001 2410000  
 WP AA199682\_24 2400001 2510000  
 WP AA199682\_25 2500001 2610000  
 WP AA199682\_26 2600001 2710000  
 WP AA199682\_27 2700001 2810000  
 WP AA199682\_28 2800001 2910000  
 WP AA199682\_29 2900001 3010000  
 WP AA199682\_30 3000001 3110000  
 WP AA199682\_31 3100001 3210000  
 WP AA199682\_32 3200001 3310000  
 WP AA199682\_33 3300001 3410000  
 WP AA199682\_34 3400001 3510000  
 WP AA199682\_35 3500001 3610000  
 WP AA199682\_36 3600001 3710000  
 WP AA199682\_37 3700001 3810000  
 WP AA199682\_38 3800001 3910000  
 WP AA199682\_39 3900001 4010000  
 WP AA199682\_40 4000001 4110000  
 WP AA199682\_41 4100001 4210000  
 WP AA199682\_42 4200001 4310000  
 WP AA199682\_43 4300001 4410000  
 WP AA199682\_44 4400001 4411529  
 ID AA199682 standard; DNA; 4411529 BP.  
 XX  
 AC AA199682;  
 XX  
 DT 15-JAN-2002 (first entry)  
 XX  
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
 XX  
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 FN US6294328-B1.  
 XX  
 PD 25-SEP-2001.  
 XX  
 PF 24-JUN-1998; 98US-00103840.  
 XX  
 PR 24-JUN-1998; 98US-00103840.  
 XX  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
 XX WPI; 2001-647261/74.  
 XX  
 DR Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 XX determining the nucleotide sequence of the strain at positions in the  
 XX genome corresponding to positions where M. tuberculosis strains CDC 1551  
 XX and H37Rv differ.  
 XX  
 PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
 XX  
 CC The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly

CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1  
 XX  
 SQ Sequence 4411529 BP; 758565A; 1449983C; 1444602G; 758379T; OU; 00Other;  
 Query Match 100.0%; Score 25; DB 4; Length 110000;  
 Best Local Similarity 100.0%; Pred. NO. 0.59;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CCGGAGGCGCAGAGGCACATAGC 25  
 Db 24426 CCGGAGGCGCAGAGGCACATAGC 24450  
 RESULT 3  
 AA199683\_00  
 WP Sequence split into 44 fragments LOCUS AA199683 Accession Aai99683  
 WP Fragment Name Begin End  
 WP AA199683\_00 1 110000  
 WP AA199683\_01 100001 210000  
 WP AA199683\_02 200001 310000  
 WP AA199683\_03 300001 410000  
 WP AA199683\_04 400001 510000  
 WP AA199683\_05 500001 610000  
 WP AA199683\_06 600001 710000  
 WP AA199683\_07 700001 810000  
 WP AA199683\_08 800001 910000  
 WP AA199683\_09 900001 1010000  
 WP AA199683\_10 1000001 1110000  
 WP AA199683\_11 1100001 1210000  
 WP AA199683\_12 1200001 1310000  
 WP AA199683\_13 1300001 1410000  
 WP AA199683\_14 1400001 1510000  
 WP AA199683\_15 1500001 1610000  
 WP AA199683\_16 1600001 1710000  
 WP AA199683\_17 1700001 1810000  
 WP AA199683\_18 1800001 1910000  
 WP AA199683\_19 1900001 2010000  
 WP AA199683\_20 2000001 2110000  
 WP AA199683\_21 2100001 2210000  
 WP AA199683\_22 2200001 2310000  
 WP AA199683\_23 2300001 2410000  
 WP AA199683\_24 2400001 2510000  
 WP AA199683\_25 2500001 2610000  
 WP AA199683\_26 2600001 2710000  
 WP AA199683\_27 2700001 2810000  
 WP AA199683\_28 2800001 2910000  
 WP AA199683\_29 2900001 3010000  
 WP AA199683\_30 3000001 3110000  
 WP AA199683\_31 3100001 3210000  
 WP AA199683\_32 3200001 3310000  
 WP AA199683\_33 3300001 3410000  
 WP AA199683\_34 3400001 3510000  
 WP AA199683\_35 3500001 3610000  
 WP AA199683\_36 3600001 3710000  
 WP AA199683\_37 3700001 3810000  
 WP AA199683\_38 3800001 3910000  
 WP AA199683\_39 3900001 4010000  
 WP AA199683\_40 4000001 4110000  
 WP AA199683\_41 4100001 4210000  
 WP AA199683\_42 4200001 4310000  
 WP AA199683\_43 4300001 4403765  
 ID AA199683 standard; DNA; 4403765 BP.  
 XX  
 AC AA199683;  
 XX  
 DT 15-JAN-2002 (first entry)  
 XX  
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
 XX  
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX  
 OS Mycobacterium tuberculosis.

```

XX  US6294328-B1.
PN
XX
XX  25-SEP-2001.
PD
XX
XX  24-JUN-1998; 98US-00103840.
PF
XX
XX  24-JUN-1998; 98US-00103840.
PR
XX
XX  (GENO-) INST GENOMIC RES.
PA
XX
XX  Fleischmann RD, White OR, Fraser CM, Venter JC;
PI
XX
XX  WPI; 2001-647261/74.
DR
XX
XX  Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT
XX  determining the nucleotide sequence of the strain at positions in the
PT
XX  genome corresponding to positions where M. tuberculosis strains CDC 1551
PT
XX  and H37Rv differ.
XX
XX  Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
PS
XX
XX  The invention relates to evaluating strain variation within and between
CC
XX  different populations of the tuberculosis bacterial pathogen,
CC
XX  Mycobacterium tuberculosis or related Mycobacterium by determining the
CC
XX  nucleotide sequence of the first strain at positions in the complete
CC
XX  sequence of the genome that correspond to positions that differ in the
CC
XX  nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and
CC
XX  H37Rv (A199682). The method is useful for evaluating strain variation of
CC
XX  M. tuberculosis and has valuable application in the fields of
CC
XX  tuberculosis genetics, epidemiology, patient treatment and epidemic
CC
XX  monitoring. Note: The sequence data for this patent did not form part of
CC
XX  the printed specification, but was obtained in electronic format directly
CC
XX  from USPTO at seqdata.uspto.gov/sequence.html?docID=6294328B1
XX
XX  SQ Sequence 4403765 BP; 757105A; 1447799C; 1441301G; 757371T; 0U; 1890ther;
SQ
XX
XX  Query Match 100.0%; Score 25; DB 4; Length 110000;
XX  Best Local Similarity 100.0%; Pred. No. 0.59;
XX  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  Qy 1 CCCGAGGGCCAGAGGGCACATAGC 25
XX  24426 CCCGAGGGCCAGAGGGCACATAGC 24450
XX
XX  RESULT 4
XX  ID ADA02684
XX  ID ADA02684 standard; DNA; 59817 BP.
XX
XX  AC ADA02684;
XX
XX  DT 06-NOV-2003 (first entry)
XX
XX  DE Human DUSP10 carcinoma associated gene, SEQ ID NO:1202.
XX
XX  KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX  prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX  gene; ds.
XX
XX  OS Homo sapiens.
XX
XX  PN WO2003057146-A2.
XX
XX  PD 17-JUL-2003.
XX
XX  PF 26-DEC-2002; 2002WO-US041414.
XX
XX  PR 26-DEC-2001; 2001US-00035832.
XX
XX  PA (SAGR-) SAGRES DISCOVERY.
XX
XX  PI Morris DW;
XX
XX  US6294328-B1.
PN
XX
XX  25-SEP-2001.
PD
XX
XX  24-JUN-1998; 98US-00103840.
PF
XX
XX  24-JUN-1998; 98US-00103840.
PR
XX
XX  (GENO-) INST GENOMIC RES.
PA
XX
XX  Fleischmann RD, White OR, Fraser CM, Venter JC;
PI
XX
XX  WPI; 2001-647261/74.
DR
XX
XX  Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT
XX  determining the nucleotide sequence of the strain at positions in the
PT
XX  genome corresponding to positions where M. tuberculosis strains CDC 1551
PT
XX  and H37Rv differ.
XX
XX  Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
PS
XX
XX  The invention relates to evaluating strain variation within and between
CC
XX  different populations of the tuberculosis bacterial pathogen,
CC
XX  Mycobacterium tuberculosis or related Mycobacterium by determining the
CC
XX  nucleotide sequence of the first strain at positions in the complete
CC
XX  sequence of the genome that correspond to positions that differ in the
CC
XX  nucleotide sequences of M. tuberculosis strains CDC 1551 (A199683) and
CC
XX  H37Rv (A199682). The method is useful for evaluating strain variation of
CC
XX  M. tuberculosis and has valuable application in the fields of
CC
XX  tuberculosis genetics, epidemiology, patient treatment and epidemic
CC
XX  monitoring. Note: The sequence data for this patent did not form part of
CC
XX  the printed specification, but was obtained in electronic format directly
CC
XX  from USPTO at seqdata.uspto.gov/sequence.html?docID=6294328B1
XX
XX  SQ Sequence 4403765 BP; 757105A; 1447799C; 1441301G; 757371T; 0U; 1890ther;
SQ
XX
XX  Query Match 100.0%; Score 25; DB 4; Length 110000;
XX  Best Local Similarity 100.0%; Pred. No. 0.59;
XX  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  Qy 1 CCCGAGGGCCAGAGGGCACATAGC 25
XX  24426 CCCGAGGGCCAGAGGGCACATAGC 24450
XX
XX  RESULT 4
XX  ID ADA02684
XX  ID ADA02684 standard; DNA; 59817 BP.
XX
XX  AC ADA02684;
XX
XX  DT 06-NOV-2003 (first entry)
XX
XX  DE Human DUSP10 carcinoma associated gene, SEQ ID NO:1202.
XX
XX  KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
XX  prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
XX  gene; ds.
XX
XX  OS Homo sapiens.
XX
XX  PN WO2003057146-A2.
XX
XX  PD 17-JUL-2003.
XX
XX  PF 26-DEC-2002; 2002WO-US041414.
XX
XX  PR 26-DEC-2001; 2001US-00035832.
XX
XX  PA (SAGR-) SAGRES DISCOVERY.
XX
XX  PI Morris DW;
XX
XX  WPI; 2003-587068/55.
PN
XX
XX  New recombinant nucleic acid encoding carcinoma associated protein,
PT
XX  useful for preparing compositions for treating carcinomas.
XX
XX  Claim 1; SEQ ID NO 1202; 245pp; English.
PS
XX
XX  The invention relates to recombinant carcinoma associated (CA) nucleic
CC
XX  acid sequences from mouse and human (ADA01482-ADA03094), and to
CC
XX  recombinant carcinoma associated proteins (CAP) encoded by them. The
CC
XX  invention also encompasses expression vectors and host cells comprising a
CC
XX  CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC
XX  binds to the protein, and a biochip comprising CA nucleic acid or
CC
XX  fragments thereof. The sequences of the invention were identified using
CC
XX  oncogenic retroviruses, which insert into the genome of the host organism
CC
XX  at random. Many of these do not carry transduced host oncogenes or
CC
XX  pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC
XX  direct consequence of the effects of proviral integration into host
CC
XX  protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC
XX  carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC
XX  leukaemia) or a propensity to carcinoma by determination of the sequence
CC
XX  of a CA gene, or by determination of CA gene expression in particular
CC
XX  tissues. CA nucleic acids, proteins and antibodies are also useful as
CC
XX  therapeutic agents and in screening and evaluating drug candidates. The
CC
XX  present sequence represents a specifically claimed human CA nucleic acid
CC
XX  sequence of the invention. Note: The complete sequence data for this
CC
XX  patent did not form part of the printed specification, but was obtained
CC
XX  in electronic format directly from WIPO at
CC
XX  ftp.wipo.int/pub/published_pct_sequences.
XX
XX  SQ Sequence 59817 BP; 16906 A; 12200 C; 12501 G; 18048 T; 0 U; 162 Other;
SQ
XX
XX  Query Match 74.4%; Score 18.6; DB 8; Length 59817;
XX  Best Local Similarity 84.0%; Pred. No. 2.8e+02;
XX  Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
XX  Qy 1 CCCGAGGGCCAGAGGGCACATAGC 25
XX  8918 CCCGAGGGCCAGAGGGCACATAGC 8942
XX
XX  RESULT 5
XX  ID ADB72422
XX  ID ADB72422 standard; DNA; 59817 BP.
XX
XX  AC ADB72422;
XX
XX  DT 04-DEC-2003 (first entry)
XX
XX  DE Human DUSP10 gene.
XX
XX  KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX  cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
XX  OS Homo sapiens.
XX
XX  PN WO2003009583-A2.
XX
XX  PD 30-JAN-2003.
XX
XX  PF 26-DEC-2001; 2001WO-US051291.
XX
XX  PR 02-MAR-2001; 2001US-00798586.
XX
XX  PR 23-OCT-2001; 2001US-00004113.
XX
XX  PR 08-NOV-2001; 2001US-00052482.
XX
XX  PR 30-NOV-2001; 2001US-00997722.
XX
XX  PR 20-DEC-2001; 2001US-00034650.
XX
XX  PA (SAGR-) SAGRES DISCOVERY.
XX
XX  PI Morris DW, Engelhard EK;
XX

```

DR WPI; 2003-239337/23.  
XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
XX  
XX Claim 1; SEQ ID NO 250; 2304pp; English.  
XX  
XX The invention relates to a novel recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the 660 sequences fully defined  
CC in the specification. A polynucleotide of the invention has cytostatic  
CC activity, and may have a use in gene therapy, or in a vaccine. The  
CC recombinant nucleic acids and polypeptides are useful for treating  
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
CC sarcomas. The present sequence represents a human gene of the invention.  
XX  
XX Sequence 59817 BP; 16906 A; 12200 C; 12501 G; 18048 T; 0 U; 162 Other;  
SQ  
Query Match 74.4%; Score 18.6; DB 9; Length 59817;  
Best Local Similarity 84.0%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CCGGAGGCGCCAGGCGGCACATAGC 25  
DB 8918 CCGGAGGCGCCAGGCGGCCTAGC 8942  
RESULT 6  
ID ABK64796/c  
XX ABK64796 standard; DNA; 14646 BP.  
AC ABK64796;  
XX  
XX 18-JUN-2002 (first entry)  
DT  
DE Human benign prostatic hyperplasia gene #691.  
XX  
XX Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.  
XX  
XX Homo sapiens.  
OS  
PN WO200212440-A2.  
XX  
PD 14-FEB-2002.  
XX  
XX 07-AUG-2001; 2001WO-US024708.  
XX  
XX 07-AUG-2000; 2000US-0223323P.  
PR  
XX 05-JUN-2001; 2001US-00873319.  
XX  
XX (GENE-) GENE LOGIC INC.  
PA (NISE ) JAPAN TOBACCO INC.  
XX  
XX Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;  
PI  
XX WPI; 2002-257476/30.  
DR  
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by  
PT detecting expression levels of one or more genes in prostate cells from  
PT patient that are differentially regulated compared to normal prostate  
PT cells.  
XX  
XX Disclosure; Page 373-377; 444pp; English.  
XX  
XX The invention relates to a method of diagnosing (I) the onset or  
XX progression of benign prostatic hyperplasia (BPH), or screening (II) for  
XX or identifying an agent that modulates the onset or progression of BPH.  
XX The method is based on changes in gene expression in BPH tissue isolated  
XX from patients exhibiting different clinical states of prostate  
XX hyperplasia as compared to normal prostate tissue. (I) comprises  
XX detecting the expression levels of one or more genes in prostate cells  
XX from the subject that are differentially regulated compared to normal  
XX prostate cells. (II) comprises preparing a first gene expression profile  
XX of BPH cells or BPH-like cell population, exposing the cells to the

CC agent, preparing a second gene expression profile of the agent exposed  
CC cells, and comparing the first and second gene expression profiles. (I)  
CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
CC for identifying an agent that modulates the onset or progression of BPH.  
CC The methods are useful to present information identifying the expression  
CC level in a tissue or cells, by comparing the expression level of genes  
CC given in the specification in the tissue or cells to the level of  
CC expression of gene in the database, and displaying the expression levels  
CC of at least one gene in the tissue or cell sample compared to the  
CC expression level in BPH. Agents using (II) are useful for treating BPH or  
CC prostate cancer. ABK64106-ABK64860 represent human benign prostatic  
CC hyperplasia gene sequences of the invention  
XX  
SQ Sequence 14646 BP; 4552 A; 2957 C; 2741 G; 4395 T; 0 U; 1 Other;  
Query Match 73.6%; Score 18.4; DB 6; Length 14646;  
Best Local Similarity 95.0%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 5 GAGGGCCAGGCGGCACATAG 24  
DB 2885 GAGGGCCAGGCGGCACATAG 2866  
RESULT 7  
ABL05531/c  
ID ABL05531 standard; cDNA; 1409 BP.  
XX  
AC ABL05531;  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11075.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR  
XX 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PB CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
XX  
XX P-PSDB; ABB61428.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 11075; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB572072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
SQ  
Sequence 1409 BP; 338 A; 375 C; 385 G; 311 T; 0 U; 0 Other;



XX 23-MAR-2001; 2001WO-US009231.  
 XX PF  
 XX PR  
 XX 23-MAR-2000; 2000US-0191637P.  
 XX PR  
 XX 11-JUL-2000; 2000US-00614150.  
 XX (PEXE ) PE CORP NY.  
 XX PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX PI  
 XX WPI; 2001-656960/75.  
 XX DR  
 XX P-PSDB; ABB61427.  
 XX PT  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX PT  
 XX Claim 1; SEQ ID NO 11072; 21bp + Sequence Listing; English.  
 XX PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX CC  
 XX SQ Sequence 3409 BP; 822 A; 842 C; 862 G; 883 T; 0 U; 0 Other;  
 Query Match 72.8%; Score 18.2; DB 4; Length 3409;  
 Best Local Similarity 87.0%; Pred. No. 3.5e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 3 CGGAGGGCCGAGGGCAGATAGC 25  
 DB 1566 CGGATGGCCGAGAGCAGATTGC 1544  
 RESULT 11  
 ADA03077  
 ID ADA03077 standard; DNA; 90442 BP.  
 XX AC  
 XX ADA03077;  
 XX DT  
 XX 06-NOV-2003 (first entry)  
 XX DE  
 XX Mouse MCG2257 carcinoma associated gene, SEQ ID NO:1595.  
 XX KW  
 XX Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
 XX prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
 XX gene; ds.  
 XX OS  
 XX Mus sp.  
 XX PN  
 XX WO2003057146-A2.  
 XX PD  
 XX 17-JUL-2003.  
 XX PF  
 XX 26-DEC-2002; 2002WO-US041414.  
 XX PR  
 XX 26-DEC-2001; 2001US-00035832.  
 XX (SAGR-) SAGRES DISCOVERY.  
 XX PA  
 XX Morris DW;  
 XX PI  
 XX WPI; 2003-587068/55.  
 XX DR  
 XX New recombinant nucleic acid encoding carcinoma associated protein,  
 PT useful for preparing compositions for treating carcinomas.  
 PT

XX Claim 1; SEQ ID NO 1595; 245pp; English.  
 XX PS  
 XX The invention relates to recombinant carcinoma associated (CA) nucleic  
 CC acid sequences from mouse and human (ADA01492-ADA03094), and to  
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
 CC invention also encompasses expression vectors and host cells comprising a  
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
 CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukaemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed murine CA nucleic acid  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX CC  
 XX SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;  
 Query Match 72.8%; Score 18.2; DB 8; Length 90442;  
 Best Local Similarity 87.0%; Pred. No. 4.2e+02;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 OY 2 CCGGAGGGCCGAGGGCAGATAG 24  
 DB 27716 CCAGAGGGCAGAGGGCAGATAG 27738  
 RESULT 12  
 ADA66361  
 ID ADA66361 standard; DNA; 90442 BP.  
 XX AC  
 XX ADA66361;  
 XX DT  
 XX 20-NOV-2003 (first entry)  
 XX DE  
 XX Mouse MCG2257 gene genomic DNA sequence.  
 XX KW  
 XX carcinoma-associated gene; CA gene; Rorc gene; MCG15938 gene; BAT1 gene;  
 XX Igkap1 gene; IGAPI gene; Zfp29 gene; hCG27579 gene; Kcnj9 gene;  
 XX Kcnj9 gene; Pp3cc gene; Pp3cc gene; MCG9110 gene; hCG27579 gene;  
 XX cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;  
 XX carcinoma-associated protein; CAP; cytostatic; gene therapy; anticancer;  
 XX vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;  
 XX prostate cancer; DNA vaccine; animal model; mouse; murine; ds; MCG2257.  
 XX OS  
 XX Mus sp.  
 XX PN  
 XX WO2003053224-A2.  
 XX PD  
 XX 03-JUL-2003.  
 XX PF  
 XX 20-DEC-2002; 2002WO-US041776.  
 XX PR  
 XX 20-DEC-2001; 2001US-00034650.  
 XX (SAGR-) SAGRES DISCOVERY.  
 XX PA  
 XX Morris DW, Engelhard EK;  
 XX PI  
 XX WPI; 2003-569168/53.  
 XX DR  
 XX Novel recombinant carcinoma-associated nucleic acid, useful for  
 PT evaluating the effect of a candidate carcinoma drug, and for diagnosing  
 PT



PT carcinoma.

XX Claim 1; Page 161-174; 229pp; English.

XX This invention relates to a novel recombinant carcinoma-associated (CA)

CC nucleic acid comprising a fully defined genomic, mRNA or coding sequences

CC of mouse Rorc gene or human RORC gene, mouse MCG15938 or human gene BAT1,

CC mouse Igkap1 gene or human Iqgap1 gene, mouse Zfp29 gene or human

CC hCG27579 gene, mouse Kcnj9 gene or human KCNJ9 gene, mouse Pp3cc gene or

CC human Pp3CC gene, mouse MCG9110 gene or human hCG27579 gene, as given in

CC the specification. CA genes are genes which are preferably expressed in

CC cancer cells, preferably lymphatic, breast, prostate or epithelial cells.

CC A compound which modifies the expression of the CA genes or bind to

CC carcinoma-associated proteins (CAP) may have cytostatic activity and the

CC sequences of the invention may enable the use of gene therapy or a

CC development of an anticancer vaccine. Therefore the invention may be

CC useful for diagnosis and treatment of carcinomas, especially lymphoma

CC carcinoma, breast cancer and prostate cancer. The CA genes may also be

CC useful as DNA vaccines and for generating animal models of carcinomas.

CC The present sequence is that of the mouse MCG2257 gene genomic DNA

CC sequence of the invention.

XX

SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 8; Length 90442;

Best Local Similarity 87.0%; Pred. No. 4.2e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCAGATAG 24

Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 13

ADBT2815

ID ADB72815 standard; DNA; 90442 BP.

XX

AC ADB72815;

XX

XX 04-DEC-2003 (first entry)

DT

DE Mouse MCG2257 gene.

XX

XX mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;

KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX

XX Mus sp.

OS

XX W02003008583-A2.

XX

XX 30-JAN-2003.

PD

XX

PF 26-DEC-2001; 2001WO-US051291.

XX

XX 02-MAR-2001; 2001US-00798586.

PR

XX 23-OCT-2001; 2001US-00004113.

PR

XX 08-NOV-2001; 2001US-00052482.

PR

XX 20-NOV-2001; 2001US-00997722.

PR

XX 20-DEC-2001; 2001US-00034650.

XX

PA (SAGR-) SAGRES DISCOVERY.

XX

XX Morris DW, Engelhard EK;

PI

XX WPI; 2003-239337/23.

DR

XX

PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,

PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX

XX Claim 1; SEQ ID NO 643; 2304pp; English.

PS

XX The invention relates to a novel recombinant nucleic acid comprising a

CC nucleotide sequence selected from any of the 660 sequences fully defined

CC in the specification. A polynucleotide of the invention has cytostatic

CC activity, and may have a use in gene therapy, or in a vaccine. The

CC recombinant nucleic acids and polypeptides are useful for treating

CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and

CC sarcomas. The present sequence represents a mouse gene of the invention.

XX

SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;

Query Match 72.8%; Score 18.2; DB 9; Length 90442;

Best Local Similarity 87.0%; Pred. No. 4.2e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCAGATAG 24

Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 14

AAS98106

ID AAS98106 standard; DNA; 492 BP.

XX

AC AAS98106;

XX

XX 12-MAR-2002 (first entry)

DT

XX

DE Human DNA for potential G protein-coupled receptor #63.

XX

KW Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;

KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;

KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;

KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;

KW depression; epilepsy; macular degeneration; lymphoma; melanoma;

KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;

KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;

KW tuberculosis; cognition disorder; memory disorder; anorexia;

KW hormonal release disorder; cardiovascular activity disorder;

KW pain perception disorder; obesity; diabetes; obesity; diabetes;

KW hyperlipidaemia; stroke; gene therapy.

XX

OS Homo sapiens.

XX

XX W0200185791-A1.

PN

XX

PD 15-NOV-2001.

XX

PF 11-MAY-2001; 2001WO-US015332.

XX

XX 11-MAY-2000; 2000US-0203217P.

PR

XX 18-MAY-2000; 2000US-0205945P.

PR

XX (LIFE-) LIFESPAN BIOSCIENCES INC.

PA

XX

PI Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;

XX

XX WPI; 2002-066595/09.

DR

XX

PT Novel G protein-coupled receptor polypeptides including galanin receptor

PT polypeptides useful for identifying modulators that are useful for

PT treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,

PT stroke.

XX

XX Claim 2; Page 81; 144pp; English.

PS

XX The invention relates to an isolated polypeptide encoded by a nucleic

CC acid molecule that is at least 80% identical to the G protein-coupled

CC (GPCR) polynucleotides included in the specification. Also included are

CC probes based on the GPCR sequences (including antisense probes), a host

CC cell comprising an expression vector comprising the GPCR sequence,

CC antibodies raised against the polypeptides, and methods of identifying

CC modulators of the polypeptides. The polypeptides are useful for

CC identifying modulator compounds which function as modulators, activators,

CC repressors, agonists or antagonists of the novel GPCR polypeptides

CC including the GAL4 polypeptide. The antibodies and nucleic acid probes as

described above can be used to detect the presence of the polypeptides and nucleic acids and are used to diagnose a variety of diseases or disorders in which GPCRs are involved e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease, depression, epilepsy, macular degeneration, lymphoma, melanoma, multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other diseases listed in the specification. The probes and antibodies are also useful for diagnosing cognitive and memory disorders, anorexia, hormonal release disorders, cardiovascular activity disorders, pain perception disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds that decrease or increase the expression of galanin receptor (GAL4) can be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is useful for treating the above mentioned disorders by gene therapy techniques. The present sequence is a novel GPCR polynucleotide of the invention

XX Sequence 492 BP; 137 A; 141 C; 128 G; 86 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 6; Length 492;  
Best Local Similarity 83.3%; Pred. No. 5.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCAGAGGGCACATAG 24  
DB 109 CCAGTGGGACAGAGGGCACAGAG 132

## RESULT 15

AAS42975

ID AAS42975 standard; cDNA; 637 BP.

XX AAS42975;

XX 18-DEC-2001 (first entry)

DE DNA encoding G protein-coupled receptor, nGPCR-2079.

XX Human; mental disorder; thyroid disease; renal failure; anorexia;  
KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;  
KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;  
KW Huntington's disease; Alzheimer's disease; viral infection;  
KW anorexia; hypotension; human immunodeficiency virus; type 2 diabetes;  
KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;  
KW nGPCR; ss.

XX Homo sapiens.

XX WO200162924-A2.

XX 30-AUG-2001.

XX 23-FEB-2001; 2001WO-US005989.

XX 24-FEB-2000; 2000US-0184602P.

XX 24-FEB-2000; 2000US-0184604P.

XX 24-FEB-2000; 2000US-0184606P.

XX 24-FEB-2000; 2000US-0184689P.

XX 24-FEB-2000; 2000US-0184690P.

XX 24-FEB-2000; 2000US-0184710P.

XX 24-FEB-2000; 2000US-0184712P.

XX 24-FEB-2000; 2000US-0184715P.

XX 24-FEB-2000; 2000US-0184716P.

XX 24-FEB-2000; 2000US-0184725P.

XX 24-FEB-2000; 2000US-0184822P.

XX (PHAA) PHARMACIA &amp; UPJOHN CO.

XX Vogeli G, Wood LS, Parodi LA, Lind P;

XX WPI; 2002-066595/09.

DR WPI: 2001-570632/64.

DR P-PSDB; AAU25665.

XX Novel nucleic acid and encoded nGPCR-x, used to screen for compounds for use in the treatment of mental disorders, such as Alzheimer's disease, or Parkinson's disease.

XX Claim 3; Page 75; 263pp; English.

XX The invention relates to novel isolated human G protein-coupled receptors (nGPCR-x). The nGPCR-x can be used for screening compounds which can be used to treat mental disorders, thyroid disease, renal failure, inflammatory conditions such as Crohn's disease, rheumatoid arthritis, autoimmune disorders, schizophrenia, migraine, stroke, dementia, depression, Parkinson's disease, Alzheimer's disease, and Huntington's disease. They may also be used for treating viral infections such as human immunodeficiency virus (HIV), type 2 diabetes, obesity, anorexia, hypotension, hypertension, thrombosis, myocardial infarction, atherosclerosis, cancer, and sexual dysfunction. AAS42927-AAS43036 represent the coding sequences of novel human G protein-coupled receptors, nGPCR-2031 to nGPCR-2140 respectively, as described in the invention

XX Sequence 637 BP; 177 A; 175 C; 164 G; 121 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 637;  
Best Local Similarity 83.3%; Pred. No. 5.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCAGAGGGCACATAG 24  
DB 182 CCAGTGGGACAGAGGGCACAGAG 205

## RESULT 16

AAS98057/c

ID AAS98057 standard; DNA; 993 BP.

XX AAS98057;

XX 12-MAR-2002 (first entry)

DE Human DNA for potential G protein-coupled receptor #15.

XX Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;  
KW Alzheimer's disease; amyotrophic lateral sclerosis; asthma;  
KW atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;  
KW chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;  
KW depression; epilepsy; macular degeneration; lymphoma; melanoma;  
KW multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;  
KW psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;  
KW tuberculosis; cognition disorder; memory disorder; anorexia;  
KW hormonal release disorder; cardiovascular activity disorder;  
KW pain perception disorder; obesity; diabetes; obesity; diabetes;  
KW hyperlipidaemia; stroke; gene therapy.

XX Homo sapiens.

XX WO200185791-A1.

XX 15-NOV-2001.

XX 11-MAY-2001; 2001WO-US015332.

XX 11-MAY-2000; 2000US-0203217P.

XX 18-MAY-2000; 2000US-0209945P.

XX (LIFE-) LIFESPAN BIOSCIENCES INC.

XX Brown JP, Miller M, Burmer G, Fabre-Suver C, Pritchard D;

XX WPI; 2002-066595/09.

PT Novel G protein-coupled receptor polypeptides including galanin receptor  
PT polypeptides useful for identifying modulators that are useful for  
PT treating Alzheimer's disease, psoriasis, melanoma, multiple sclerosis,  
PT stroke.  
PS Disclosure; Page 121-122; 144pp; English.  
XX  
XX The invention relates to an isolated polypeptide encoded by a nucleic  
CC acid molecule that is at least 80% identical to the G protein-coupled  
CC (GPCR) polynucleotides included in the specification. Also included are  
CC probes based on the GPCR sequences (including antisense probes), a host  
CC cell comprising an expression vector comprising the GPCR sequence,  
CC antibodies raised against the polypeptides, and methods of identifying  
CC modulators of the polypeptides. The polypeptides are useful for  
CC identifying modulator compounds which function as modulators, activators,  
CC repressors, agonists or antagonists of the novel GPCR polypeptides  
CC including the GAL4 polypeptide. The antibodies and nucleic acid probes as  
CC described above can be used to detect the presence of the polypeptides  
CC and nucleic acids and are used to diagnose a variety of diseases or  
CC disorders in which GPCRs are involved e.g., Alzheimer's disease,  
CC amyotrophic lateral sclerosis, asthma, atherosclerosis, basal cell  
CC carcinoma, breast carcinoma, cardiomyopathy, chondrosarcoma, chronic  
CC obstructive pulmonary disease, Crohn's disease, depression, epilepsy,  
CC osteoarthritis, osteoporosis, Parkinson's disease, psoriasis, rheumatoid  
CC arthritis, schizophrenia, ulcerative colitis, tuberculosis and many other  
CC diseases listed in the specification. The probes and antibodies are also  
CC useful for diagnosing cognition and memory disorders, anorexia, hormonal  
CC release disorders, cardiovascular activity disorders, pain perception  
CC disorders, obesity, diabetes, Alzheimer's disease. Preferably, compounds  
CC that decrease or increase the expression of galanin receptor (GAL4) can  
CC be used to treat obesity, diabetes, hyperlipidaemia and stroke. The GPCR  
CC nucleic acid is useful for treating the above mentioned disorders by gene  
CC therapy techniques. The present sequence is a novel GPCR polynucleotide  
CC of the invention  
XX  
XX Sequence 993 BP; 213 A; 261 C; 248 G; 271 T; 0 U; 0 Other;  
SQ  
Query Match 70.4%; Score 17.6; DB 6; Length 993;  
Best Local Similarity 83.3%; Pred. No. 5.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CCCGAGGGCCAGAGGGGCACATAG 24  
DB 648 CCAGGTGGGACAGAGGGGCACAG 625  
RESULT 17  
ADC86886/c  
ID ADC86886 standard; DNA; 1152 BP.  
XX  
XX AC ADC86886;  
XX  
XX DT 01-JAN-2004 (first entry)  
XX  
XX DE Human GPCR gene SEQ ID NO:1339.  
XX  
XX ds; gene; human; GPCR;  
XX guanosine triphosphate-binding protein coupled receptor; gene therapy.  
XX Homo sapiens.  
XX  
XX OS EP1270724-A2.  
XX  
XX PD 02-JAN-2003.  
XX  
XX PF 18-JUN-2002; 2002EP-00013517.  
XX  
XX PR 18-JUN-2001; 2001JP-00246789.  
XX  
XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
XX  
PI Suwa M, Asai K, Akiyama Y, Aburatani H;  
XX WPI; 2003-315783/31.  
DR P-PSDB; ADC86887.  
XX  
XX PT New polynucleotide, useful for preparing a composition for treating a  
PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.  
XX  
XX PS Claim 1; SEQ ID NO 1339; 28pp; English.  
XX  
XX CC The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient  
CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The  
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
CC invention.  
XX  
XX SQ Sequence 1152 BP; 221 A; 303 C; 307 G; 321 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 9; Length 1152;  
Best Local Similarity 83.3%; Pred. No. 5.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CCCGAGGGCCAGAGGGGCACATAG 24  
DB 554 CCAGGTGGGACAGAGGGGCACAG 531  
RESULT 18  
ADE28123  
ID ADE28123 standard; cDNA; 3311 BP.  
XX  
XX AC ADE28123;  
XX  
XX DT 29-JAN-2004 (first entry)  
XX  
XX DE Human NTRAN cDNA - SEQ ID 28.  
XX  
XX KW human; neurotransmission-associated protein; NTRAN; cytostatic;  
XX immunomodulator; immune disorder; cancer; gene therapy; ss; gene.  
XX OS Homo sapiens.  
XX  
XX PN WO2003051902-A1.  
XX  
XX PD 26-JUN-2003.  
XX  
XX PF 12-DEC-2002; 2002WO-US040059.  
XX  
XX PR 14-DEC-2001; 2001US-0340798P.  
XX  
XX PR 18-MAR-2002; 2002US-0365645P.  
XX  
XX PR 25-MAR-2002; 2002US-0367662P.  
XX  
XX PR 10-MAY-2002; 2002US-0379887P.  
XX  
XX PR 31-MAY-2002; 2002US-0384639P.  
XX  
XX PA (INCY-) INCYTE GENOMICS INC.  
XX  
XX PI Baughn MR, Bhatia U, Blake JJ, Burrill JD, Elliott VS;  
PI Emerling BM, Forsythe IJ, Gietzen KJ, Gorvad AE, Griffin JA;  
PI Hafalia AJA, Ho A, Jackson AA, Jiang X, Kable AE, Kearney L;  
PI Khare R, Lee EA, Lee S, Lu DAM, Marquis JP, Lehn-Nelson PW;  
PI Rankumar J, Richardson TW, Sprague WW, Tran UK, Chawla NK;  
PI Warren BA, Yue H, Zheng W;  
XX  
XX WPI; 2003-514037/48.  
DR P-PSDB; ADE28101.  
XX  
XX New human neurotransmission-associated proteins (NTRAN) polypeptide,  
PT useful for preparing a composition for treating a disease associated with  
PT decreased expression or overexpression of NTRAN e.g., cancer.

XX Claim 5; SEQ ID NO 28; 261bp; English.  
XX The invention relates to a novel isolated human neurotrophin-  
CC associated proteins (NTRN) polypeptide. The polypeptide of the invention  
CC demonstrates cytosolic and immunomodulator activities and may be useful  
CC for preparing a composition for diagnosing or treating a disease or  
CC condition associated with decreased expression or overexpression of  
CC functional NTRN including immune disorders or cancer, as well as during  
CC gene therapy procedures. The current sequence is that of the human NTRN  
CC cDNA of the invention.  
XX  
SQ Sequence 3311 BP; 676 A; 1038 C; 1058 G; 539 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 9; Length 3311;  
Best Local Similarity 83.3%; Pred. No. 6.2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 CCGAGGGCCAGGACATAGC 25  
DB 81 CCGAGGGCCAGGACATAGC 104  
RESULT 19  
AAK88002/c  
ID AAK88002 standard; cDNA; 373 BP.  
XX AC AAK88002;  
XX  
XX 05-NOV-2001 (first entry)  
DE Human digestive system antigen coding sequence SEQ ID NO: 318.  
XX  
XX Human digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ss.  
XX Homo sapiens.  
OS  
XX  
XX WO200155314-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001324.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198122P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 03-SEP-2000; 2000US-0229509P.  
PR 03-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249285P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 11-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-502630/55.  
DR P-PSDB; RAM92229.  
XX  
XX polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases.  
XX  
XX Claim 1; SEQ ID NO 318; 986pp; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
CC system antigen of the invention  
XX  
SQ Sequence 373 BP; 71 A; 107 C; 101 G; 91 T; 0 U; 3 Other;  
Query Match 68.8%; Score 17.2; DB 4; Length 373;  
Best Local Similarity 86.4%; Pred. No. 8.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 4 GGAGGGCCAGAGGGGACATAGC 25  
DB 297 GGAGGGCCAGAGGGGACACAGC 276  
RESULT 20  
ADA73174  
ID ADA73174 standard; DNA; 2000 BP.  
XX  
XX ADA73174;  
AC

XX 20-NOV-2003 (first entry)  
XX Rice gene, SEQ ID 6500.  
DE  
XX Plant; bacterial infection; fungal infection; viral infection; rice;  
XX gene, ds.  
OS Oryza sativa.  
XX WO2003000898-A1.  
PN  
XX 03-JAN-2003.  
PD  
XX 22-JUN-2001; 2001WO-IB001105.  
PF  
XX 22-JUN-2001; 2001WO-IB001105.  
PR  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
PA  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
PI  
XX WPI; 2003-175290/17.  
DR  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
XX Claim 27; SEQ ID NO 6500; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX  
SQ Sequence 2000 BP; 656 A; 391 C; 358 G; 593 T; 0 U; 2 Other;  
Query Match 68.8%; Score 17.2; DB 7; Length 2000;  
Best Local Similarity 86.4%; Pred. No. 8.9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
OY 2 CCGAGGGCCAGAGGGGACACATA 23  
DB 1819 CCGAGGGCCCGTGGCACATA 1840  
RESULT 21  
AAK89548/c  
ID AAK89548 standard; DNA; 2266 BP.  
XX  
XX AAK89548;  
AC  
XX  
XX 05-NOV-2001 (first entry)  
DT  
XX Human digestive system antigen genomic sequence SEQ ID NO: 3124.  
DE  
XX Human, digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200155314-A2.  
PN  
XX 02-AUG-2001.  
PD

XX  
 PF 17-JAN-2001; 2001WC-US001324.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184564P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189974P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-MAR-2000; 2000US-0198123P.  
 PR 19-MAR-2000; 2000US-0200515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0215647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0228272P.  
 PR 22-AUG-2000; 2000US-0228681P.  
 PR 22-AUG-2000; 2000US-0228682P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0228287P.  
 PR 01-SEP-2000; 2000US-0228343P.  
 PR 01-SEP-2000; 2000US-0228344P.  
 PR 01-SEP-2000; 2000US-0228345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231245P.  
 PR 08-SEP-2000; 2000US-0231246P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234224P.  
 PR 25-SEP-2000; 2000US-0234297P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.

PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0241826P.  
 PR 08-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250351P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 11-DEC-2000; 2000US-0251990P.  
 PR 05-JAN-2001; 2001US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-502630/55.

XX

PT Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases.  
XX  
XX  
PS Disclosure; SEQ ID NO 3124; 986pp; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
CC encoding a digestive system antigen of the invention  
XX  
XX Sequence 2266 BP; 486 A; 594 C; 578 G; 608 T; 0 U; 0 Other;  
SQ  
Query Match 68.8%; Score 17.2; DB 4; Length 2266;  
Best Local Similarity 86.4%; Pred. No. 9e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 GGAGGGCCAGAGGCACATAGC 25  
DB 712 GGAGGGCCAGAGGGGACACAGC 691  
RESULT 22  
ABI99772/C  
ID ABI99772 standard; cDNA; 10733 BP.  
XX  
XX AC ABI99772;  
XX  
XX 07-MAR-2002 (first entry)  
XX  
XX Mouse ischaemic condition related cDNA sequence SEQ ID NO:859.  
XX  
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
XX  
XX Mus musculus.  
XX  
XX WO200198188-A2.  
XX  
XX 22-NOV-2001.  
XX  
XX 18-MAY-2001; 2001WO-JP004192.  
XX  
XX 18-MAY-2000; 2000JP-00145977.  
XX  
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
XX  
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
XX  
XX WPI; 2002-034733/04.  
XX  
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
XX expression levels of particular genes defined in the specification or by  
XX determining the expression profile of a gene group comprising these  
XX genes.  
XX  
XX Claim 2; Page 2137-2143; 2690pp; English.  
XX  
XX The present invention describes a method for examining ischaemic  
XX conditions, comprising measuring the expression levels of particular  
XX genes (i) in a test sample or determining the expression profile of a  
XX gene group in the sample comprising genes selected from (i). The method  
XX is useful for examining the ischaemic condition (e.g. compressive  
XX ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
XX expression levels of particular genes (ABI99202 to ABI99912, encoding the  
XX protein sequences in AB57020 to AB57374) or by determining the  
XX expression profile of a gene group comprising these genes. The expression  
XX levels or expression profiles produced by these genes are used as an  
XX indicator when screening for ischaemic condition-improving drugs or  
XX therapeutics for ischaemic diseases. ABI99913 and ABI99914 represent PCR

CC primers for a mouse ischaemic condition related sequence, which are used  
CC in the exemplification of the present invention  
XX  
XX Sequence 10733 BP; 2166 A; 2890 C; 2968 G; 2704 T; 0 U; 5 Other;  
SQ  
Query Match 68.8%; Score 17.2; DB 6; Length 10733;  
Best Local Similarity 86.4%; Pred. No. 9.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 3 CGAGGGCCAGAGGCACATAG 24  
DB 6259 CGAGGGCCAGAGGCACAGAG 6237  
RESULT 23  
AAK89507/C  
ID AAK89507 standard; DNA; 16038 BP.  
XX  
XX AC AAK89507;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human digestive system antigen genomic sequence SEQ ID NO: 3083.  
XX  
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
XX digestive system disorder; Meckel's diverticulum; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155314-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001324.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180828P.  
XX 24-FEB-2000; 2000US-0184564P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 07-JUL-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226686P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.







XX (SAGR-) SAGRES DISCOVERY.  
 XX Morris DW, Engelhard EK;  
 XX WPI; 2003-239337/23.  
 XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.  
 XX Claim 1; SEQ ID NO 73; 2304pp; English.  
 XX The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a mouse gene of the invention.  
 XX Sequence 23070 BP; 5504 A; 5904 C; 5830 G; 5559 T; 0 U; 273 Other;  
 SQ

Query Match 68.8%; Score 17.2; DB 9; Length 23070;  
 Best Local Similarity 86.4%; Pred. No. 1e+03;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGGAGGCGCAGGGGCACATA 23  
 DB 18715 CTGGAGTGCTAGGGGCACATA 18736

RESULT 27  
 ID AAF18390/c  
 XX AAF18390 standard; DNA; 376 BP.  
 XX AAF18390;  
 XX 14-MAR-2001 (first entry)  
 DT Lung cancer associated polynucleotide sequence SEQ ID 409.  
 DE Human; lung cancer associated protein; neuroprotective; cytostatic;  
 KW Cardioactive; immunomodulatory; muscular active; vulnerary;  
 KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease; ds.  
 XX Homo sapiens.  
 OS WO200055180-A2.  
 XX 21-SEP-2000.  
 XX 08-MAR-2000; 2000WO-US005918.  
 XX 12-MAR-1999; 99US-0124270P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX Ruben SM;  
 PI WPI; 2000-587514/55.  
 DR P-PSDB; AAB5814.  
 XX Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer.  
 XX Claim 1; Page 877; 1425pp; English.  
 PS Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer

CC associated proteins and polynucleotide sequences, their agonists, and  
 CC antagonists may have neuroprotective; cytostatic; cardioactive;  
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
 CC general; nephrotropic; antiinfective; gynecological; or antibacterial  
 CC activity. The invention also includes antibodies specific for the protein  
 CC or polynucleotide sequences. The lung cancer associated polynucleotide  
 CC sequences may be used for detection of lung cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The proteins may be used to treat disorders such as  
 CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,  
 CC cardiovascular, renal, and proliferative disorders. The proteins may also  
 CC be used in the treatment of wounds and infectious diseases.  
 CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are  
 CC used in the course of the invention for the identification and  
 CC characterisation of the polynucleotide and protein sequences  
 XX

SQ Sequence 376 BP; 78 A; 93 C; 118 G; 77 T; 0 U; 10 Other;  
 Query Match 68.0%; Score 17; DB 3; Length 376;  
 Best Local Similarity 100.0%; Pred. No. 9.9e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GAGGGCCAGAGGGCACA 21  
 DB 143 GAGGGCCAGAGGGCACA 127

RESULT 28  
 ID ACH19776/c  
 XX ACH19776 standard; cDNA; 406 BP.  
 XX ACH19776;  
 XX 13-OCT-2003 (first entry)  
 DT Human adult lung cDNA #779.  
 DE Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
 KW genome mapping; biodiversity; genetic disorder.  
 XX Homo sapiens.  
 OS US2003073623-A1.  
 XX 17-APR-2003.  
 XX 30-JUL-2001; 2001US-00918995.  
 XX 30-JUL-2001; 2001US-00918995.  
 XX (DRMA/) DRMANAC R T.  
 PA (LABA/) LABAT I.  
 PA (STAC/) STACHE-CRAIN B.  
 PA (DICK/) DICKSON M C.  
 PA (JONE/) JONES L W.  
 XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
 XX WPI; 2003-615964/58.  
 XX New polynucleotide sequences obtained from various cDNA libraries, useful  
 PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
 PT mapping, in the recombinant production of protein, or in generating  
 PT antisense DNA or RNA.  
 XX Claim 1; SEQ ID NO 6988; 44pp; English.  
 PS The invention relates to an isolated polynucleotide comprising any one of  
 CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was  
 CC determined by the technique of SBH (sequencing by hybridisation). Also  
 CC included is a purified polypeptide comprising a sequence corresponding to  
 CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
 CC are useful in diagnostics as expressed sequence tags (EST) for

CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC [seqdata.uspto.gov/sequence.html?DocID=20030073623](http://seqdata.uspto.gov/sequence.html?DocID=20030073623)  
XX  
SQ Sequence 406 BP; 80 A; 116 C; 132 G; 78 T; 0 U; 0 Other;

Query Match 68.0%; Score 17; DB 8; Length 406;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GAGGCCAGAGGGCACA 21  
|||||  
Db 73 GAGGCCAGAGGGCACA 57

RESULT 29.  
AAF18383/c  
ID AAF18383 standard; DNA; 437 BP.

XX AC AAF18383;

DT 14-MAR-2001 (first entry)

DE Lung cancer associated polynucleotide sequence SEQ ID 402.

XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
KW cardiactive; immunomodulatory; muscular active; vulnerary;  
KW gastrointestinal; nephrotropic; antiinfective; gynecological;  
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KW proliferative disorder; wound healing; infectious disease; ds.

XX Homo sapiens.

XX WO200055180-A2.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US005918.

XX 12-MAR-1999; 99US-0124270P.

XX (HUMA-) HUMAN GENOME SCI INC.  
XX (ROSE/) ROSEN C A.

XX Ruben SM;

XX WPI; 2000-587514/55.

XX P-PSDB; AAB59507.

XX Lung cancer associated gene sequences, referred to as lung cancer  
XX antigens, useful for treatment, prevention, and diagnosis of disorders  
XX such as lung cancer.

XX Claim 1; Page 868; 1425pp; English.

XX Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
XX associated proteins represented in AAB59106 - AAB59548. Lung cancer  
XX associated proteins and polynucleotide sequences, their agonists, and  
XX antagonists may have neuroprotective; cytostatic; cardioactive;  
XX immunomodulatory; muscular active general; vulnerary; gastrointestinal  
XX general; nephrotropic; antiinfective; gynecological; or antibacterial  
XX activity. The invention also includes antibodies specific for the protein  
XX or polynucleotide sequences. The lung cancer associated polynucleotide  
XX sequences may be used for detection of lung cancer, chromosome  
XX identification, as chromosome markers, and for numerous other diagnostic

CC or research purposes. The proteins may be used to treat disorders such as  
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,  
CC cardiovascular, renal, and proliferative disorders. The proteins may also  
CC be used in the treatment of wounds and infectious diseases.  
CC Polynucleotide sequences AAF18425 - AAF18433 and peptide AAB58549 are  
CC used in the course of the invention for the identification and  
CC characterisation of the polynucleotide and protein sequences

XX Sequence 437 BP; 87 A; 126 C; 142 G; 81 T; 0 U; 1 Other;

Query Match 68.0%; Score 17; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 1e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GAGGCCAGAGGGCACA 21  
|||||  
Db 75 GAGGCCAGAGGGCACA 59

RESULT 30  
ABL88406/c  
ID ABL88406 standard; cDNA; 547 BP.

XX AC ABL88406;

DT 16-MAY-2002 (first entry)

DE Pain regulated cDNA sequence 49.

KW Pain; analgesic; gene therapy; neurological disorder;  
KW neurodegenerative disease; Gene; ss.

XX Rattus sp.

XX WO200212338-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-EP009011.

XX 03-AUG-2000; 2000DE-01037759.

XX (CHEF ) GRUENTHAL GMBH.

XX Gillen C, Wetzels I, Wrenndt S, Weihe E, Schaefer MK;

XX WPI; 2002-257459/30.

XX Identifying pain-regulating compounds, useful for treating chronic pain  
XX and for diagnosis, by measuring binding of compounds to specific peptides  
XX and proteins.

XX Claim 14; Fig 29; 213pp; German.

XX The invention relates to identifying pain-regulating substances (A)  
XX comprises (i) incubating a test substance with a cell (or preparation  
XX from it) that has synthesised a peptide or protein (B) and (ii) measuring  
XX either binding of the test substance to (B) or some functional parameter  
XX that is altered by this binding. The method is useful for identifying  
XX pain-regulating substances (A) with analgesic activity. (A) along with  
XX nucleic acid (ABL88411-ABL88441) that encode proteins (B, ABB85006-  
XX ABB85037) that interact with (A); (B); vectors containing the nucleic  
XX acid; antibodies against (B); cells that express (B) and agents that bind  
XX to (B), are all useful for treating pain, particularly chronic pain,  
XX including use in gene therapy. The same materials can also be used for  
XX diagnosis, e.g. of neurological and neurodegenerative diseases. The  
XX present sequence is that of a polynucleotide of the invention

XX Sequence 547 BP; 145 A; 151 C; 91 G; 158 T; 0 U; 2 Other;

Query Match 68.0%; Score 17; DB 6; Length 547;  
Best Local Similarity 80.0%; Pred. No. 1e+03;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Mon Jun 21 09:02:13 2004

us-10-624-714-17.rng

Page 20

Qy 1 CCCGAGGGCCAGAGGGACATAGC 25  
386 CCCGAAAGGAAGGACACATAGC 362  
Db

Search completed: June 20, 2004, 10:18:06  
Job time : 160.776 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1168.25 Seconds  
(without alignments)  
639.034 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccggaggccagggcacatagc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estini:\*

4: em\_estini:\*

5: em\_estov:\*

6: em\_estov:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gssl:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.2	76.8	675	10	BE309187
C 2	19.2	76.8	861	14	CF242277
C 3	18.8	75.2	452	12	BJ036626
C 4	18.8	75.2	637	10	AW645104
					-----
					BE309187 601095072
					CF242277 AGENCOURT
					BJ036626 BJ036626
					AW645104 Cn48h12.w

AZ189837	SP 1015 A	760	28	AZ189837	75.2	18.8
BH312135	CH230-101	768	28	BH312135	75.2	18.8
BF248203	601821202	866	10	BF248203	75.2	18.8
CB587782	AGENCOURT	866	14	CB587782	75.2	18.8
CD215458	PPpnr.pk0	236	14	CD215458	74.4	18.6
ADZ76396	RPCI-24-1	303	28	ADZ76396	74.4	18.6
BY124227	BY124227	326	13	BY124227	74.4	18.6
BY008855	BY008855	351	13	BY008855	74.4	18.6
BB800951	BB800951	358	10	BB800951	74.4	18.6
BM540159	hb18g12.9	584	12	BM540159	74.4	18.6
BM538550	ha9ra10.9	625	12	BM538550	74.4	18.6
BB657042	BB657042	637	10	BB657042	74.4	18.6
BI652362	603299633	651	12	BI652362	74.4	18.6
BF215493	601881056	734	10	BF215493	74.4	18.6
AL164025	Tetraodon	734	29	CNS013K0	74.4	18.6
BI109078	602896924	825	12	BI109078	74.4	18.6
BI646763	603279793	839	12	BI646763	74.4	18.6
AL13187	Tetraodon	945	29	CNS02THM	74.4	18.6
BE887139	601508337	1050	10	BE887139	74.4	18.6
AK051753	Mus muscu	3251	11	AK051753	74.4	18.6
BZ344555	hg87c01.b	350	28	BZ344555	73.6	18.4
CE247615	tigr-gss-	587	29	CE247615	73.6	18.4
AL565528	AL565528	1201	9	AL565528	73.6	18.4
BA483005	BA483005	275	10	BA483005	72.8	18.2
BY556125	BY556125	446	13	BY556125	72.8	18.2
BI243418	RE41056.5	466	12	BI243418	72.8	18.2
AI106672	GH06745.5	478	9	AI106672	72.8	18.2
BI167064	RE07351.5	503	12	BI167064	72.8	18.2
BZ860665	CH240 256	527	28	BZ860665	72.8	18.2
AZ478773	LM0295K06	541	28	AZ478773	72.8	18.2
AI135996	GH13981.5	576	9	AI135996	72.8	18.2
AI108894	GH08206.5	578	9	AI108894	72.8	18.2
AI544125	GH27186.5	578	9	AI544125	72.8	18.2
AI106672	GH06745.5	583	9	AI106672	72.8	18.2
AI106827	GH06061.5	586	9	AI106827	72.8	18.2
AI063819	GH03910.5	588	9	AI063819	72.8	18.2
AI258372	LF01552.5	594	9	AI258372	72.8	18.2
BI589463	RH23274.5	595	12	BI589463	72.8	18.2
AQ763553	HS_3156_E	601	28	AQ763553	72.8	18.2
BG636173	SD13773.5	602	12	BG636173	72.8	18.2
AI064563	GH05023.5	611	9	AI064563	72.8	18.2
AI133912	GH10955.5	614	9	AI133912	72.8	18.2
BI376192	RE62049.5	619	12	BI376192	72.8	18.2
BI354430	GM16474.5	629	12	BI354430	72.8	18.2
BI168276	RE08877.5	630	12	BI168276	72.8	18.2
BI358748	RE37095.5	640	12	BI358748	72.8	18.2
CE734784	tigr-gss-	640	29	CE734784	72.8	18.2
AI405779	GH25908.5	643	9	AI405779	72.8	18.2
AI135368	GH13067.5	648	9	AI135368	72.8	18.2
BI642035	SD24007.5	652	12	BI642035	72.8	18.2
AI064545	GH05002.5	657	9	AI064545	72.8	18.2
BI487260	RE07026.5	657	12	BI487260	72.8	18.2
CE417348	tigr-gss-	701	29	CE417348	72.8	18.2
CNS01HHJ	ANopheles	733	29	CNS01HHJ	72.8	18.2
AA803450	GM13207.5	748	9	AA803450	72.8	18.2
AL293291	Tetraodon	869	29	CNS04UAQ	72.8	18.2
BF315659	601899312	976	10	BF315659	72.8	18.2
AL145976	ANopheles	1007	29	CNS01LNR	72.8	18.2
BI103079	EX103079	343	13	BI103079	72.8	18.2
BI122652	Danio rer	527	29	BI122652	72.0	17.8
AL977485	Danio rer	672	29	AL977485	72.0	17.8
BX203153	Danio rer	745	29	BX203153	72.0	17.8
BI182875	Danio rer	800	29	BI182875	72.0	17.8
BF234179	602026039	894	10	BF234179	72.0	17.8
BI381668	EX381668	1190	13	BI381668	72.0	17.8
AW855329	FW1-CT026	237	10	AW855329	71.2	17.8
CE197537	tigr-gss-	263	29	CE197537	71.2	17.8
BI685011	603307683	395	12	BI685011	71.2	17.8
CG384508	OGVGA15TV	425	29	CG384508	71.2	17.8
AL747614	Danio rer	486	29	DR12J2S	71.2	17.8
BI772144	603054355	491	12	BI772144	71.2	17.8
CC704797	OGVBM53TV	491	29	CC704797	71.2	17.8
CC682864	OGUSL58TV	495	29	CC682864	71.2	17.8

78 17.8 71.2 506 10 B8638537 946010D06  
 79 17.8 71.2 549 28 AQ529507 RPCI-11-3  
 80 17.8 71.2 593 29 CC692742 CGUHE59TV  
 81 17.8 71.2 613 29 BX126344 Danio rer  
 82 17.8 71.2 646 29 BX226140 Danio rer  
 83 17.8 71.2 648 29 BX125831 Danio rer  
 84 17.8 71.2 655 29 BX162646 Danio rer  
 85 17.8 71.2 661 29 BX230269 Danio rer  
 86 17.8 71.2 709 28 BX224716 Danio rer  
 87 17.8 71.2 722 28 BX224716 CH230-419  
 88 17.8 71.2 731 29 BX147008 Danio rer  
 89 17.8 71.2 737 14 CA604722 wrl.pK004  
 90 17.8 71.2 745 29 BX160297 Danio rer  
 91 17.8 71.2 771 29 BX189398 Danio rer  
 92 17.8 71.2 774 29 BX161442 Danio rer  
 93 17.8 71.2 845 28 B2835651 B2835651 CH240-209  
 94 17.8 71.2 866 12 BG445381 GA\_EA002  
 95 17.8 71.2 1128 10 BF672836 602152675  
 96 17.8 71.2 1201 13 BX353858 BX353858  
 97 17.8 71.2 1201 13 BX402018 BX402018  
 98 17.8 71.2 185 28 A2466655 A2466655 IM0277E03  
 99 17.8 71.2 260 10 AW426057 AW426057 59426 MAR  
 100 17.8 71.2 275 9 AW783914 c8h11a1.f  
 101 17.8 71.2 294 12 BG401235 BG401235 602465333  
 102 17.8 71.2 296 10 BX194338 BX194338  
 103 17.8 71.2 298 29 C3368901 C3368901 tigr-gss-  
 104 17.8 71.2 343 9 AA555977 AA555977 v845b09.r  
 105 17.8 71.2 359 10 B8685008 B8685008  
 106 17.8 71.2 364 28 AQ235341 AQ235341 HS\_2055.A  
 107 17.8 71.2 403 10 B8689841 B8689841  
 108 17.8 71.2 405 10 B9463110 UI-M-BZ0-  
 109 17.8 71.2 408 10 B9693108 B9693108  
 110 17.8 71.2 412 9 AI454036 AI454036 v845b09.y  
 111 17.8 71.2 414 10 B9552982 B9552982 UI-M-BH1-  
 112 17.8 71.2 418 13 BX637670 BX637670  
 113 17.8 71.2 420 10 AW120563 AW120563 UI-M-BH1-  
 114 17.8 71.2 428 10 AW124039 AW124039 UI-M-BH2-  
 115 17.8 71.2 437 10 B3684219 B3684219  
 116 17.8 71.2 440 9 AW060424 AW060424 UI-M-BH1-  
 117 17.8 71.2 452 10 B8157245 B8157245 RC4-HT037  
 118 17.8 71.2 456 10 AW909327 AW909327 uf47d01.x  
 119 17.8 71.2 468 10 BF011846 BF011846 us42501.y  
 120 17.8 71.2 472 10 BF776862 BF776862 288604.MA  
 121 17.8 71.2 476 10 B8769410 B8769410  
 122 17.8 71.2 494 12 B8689205 B8689205 337113.BA  
 123 17.8 71.2 495 9 AV111570 AV111570  
 124 17.8 71.2 499 12 BW921509 BW921509 AGENCOURT  
 125 17.8 71.2 510 25 P487L AL160767 Leishmani  
 126 17.8 71.2 528 14 C8440511 C8440511 690800.MA  
 127 17.8 71.2 529 9 AV396810 AV396810  
 128 17.8 71.2 540 13 BX525039 BX525039  
 129 17.8 71.2 544 28 B2103023 CH230-147  
 130 17.8 71.2 546 10 BF041030 BF041030 BP250018B  
 131 17.8 71.2 550 9 AI791866 AI791866 oq93807.y  
 132 17.8 71.2 552 14 C8429212 C8429212 605013.MA  
 133 17.8 71.2 626 10 B8743077 B8743077 601575132  
 134 17.8 71.2 641 29 C8201635 C8201635 tigr-gss-  
 135 17.8 71.2 649 14 C8443388 C8443388 694266.MA  
 136 17.8 71.2 652 29 C8134385 C8134385 tigr-gss-  
 137 17.8 71.2 656 14 C8245965 C8245965 UI-M-F00-  
 138 17.8 71.2 677 14 C8753584 C8753584 002200190  
 139 17.8 71.2 804 28 C8375846 C8375846 PURGC04TD  
 140 17.8 71.2 811 28 B2216631 B2216631 CH230-281  
 141 17.8 71.2 812 12 B8611061 B8611061 602612005  
 142 17.8 71.2 834 29 C80305M5 AL182246 Tetraodon  
 143 17.8 71.2 853 10 B8036253 B8036253 MO21E11.M  
 144 17.8 71.2 860 28 B8547678 B8547678 601076264  
 145 17.8 71.2 860 28 B2805951 B2805951 PURFK30TD  
 146 17.8 71.2 861 28 B2805946 B2805946 PURFK30TB  
 147 17.8 71.2 897 14 C8984091 C8984091  
 148 17.8 71.2 909 13 B8642455 B8642455 AGENCOURT  
 149 17.8 71.2 915 13 B8439446 B8439446 AGENCOURT  
 150 17.8 71.2 935 10 BF308856 BF308856 601889783

151 17.6 70.4 938 13 BU432132 BU432132  
 152 17.6 70.4 981 29 C8S04NL9 C8S04NL9  
 153 17.6 70.4 984 29 CQ295025 CQ295025  
 154 17.6 70.4 1024 29 C8S05CR5 C8S05CR5  
 155 17.6 70.4 1025 28 B2179185 B2179185  
 156 17.6 70.4 1043 29 C8S05S74 C8S05S74  
 157 17.6 70.4 1047 29 C8S05A0B C8S05A0B  
 158 17.6 70.4 1121 14 C8S83237 C8S83237  
 159 17.6 70.4 1126 12 B8S62134 B8S62134  
 160 17.6 70.4 126 11 AK075847 AK075847  
 161 17.6 70.4 1731 10 BF217741 BF217741 601882568  
 162 17.6 70.4 237 10 B2022272 B2022272  
 163 17.6 70.4 297 10 BF606195 BF606195 273167.MA  
 164 17.6 70.4 441 14 C8541304 C8541304  
 165 17.6 70.4 463 28 AQ885822 AQ885822 HS\_5527.B  
 166 17.6 70.4 481 14 CB728441 CB728441  
 167 17.6 70.4 583 28 A2995132 A2995132 2M0280A20  
 168 17.6 70.4 603 14 C8541739 C8541739 C0609H03-  
 169 17.6 70.4 624 14 C8S78383 C8S78383 AMGNNUC-N  
 170 17.6 70.4 672 29 AG135212 AG135212 Pan.trog1  
 171 17.6 70.4 718 14 C8S36153 C8S36153 UI-M-G10-  
 172 17.6 70.4 743 29 BX136993 BX136993 Danio rer  
 173 17.6 70.4 759 12 BM010948 BM010948 603634592  
 174 17.6 70.4 808 29 C8994191 C8994191 Z0AAX73TH  
 175 17.6 70.4 838 29 CG121728 CG121728 PUFTH87TB  
 176 17.6 70.4 858 29 CG088296 CG088296 PUFTH87TB  
 177 17.6 70.4 904 29 CG121729 CG121729 PUFTH87TD  
 178 17.6 70.4 924 13 BU147085 BU147085 AGENCOURT  
 179 17.6 70.4 935 10 B8970642 B8970642 601679871  
 180 17.6 70.4 960 14 C8S72068 C8S72068 AL532208  
 181 17.6 70.4 1101 29 C8S00H2A C8S00H2A  
 182 17.6 70.4 1101 11 AK040162 AK040162 Drosophil  
 183 17.6 70.4 141 29 C8586171 C8586171 CH240-383  
 184 17.6 70.4 141 29 C8343037 C8343037 tigr-gss-  
 185 17.6 70.4 220 12 B8287804 B8287804 526609.MA  
 186 17.6 70.4 280 9 AV330466 AV330466 AV330466  
 187 17.6 70.4 280 10 B8478724 B8478724  
 188 17.6 70.4 282 9 AL921397 AL921397  
 189 17.6 70.4 287 12 B1740718 B1740718 GC84e08.y  
 190 17.6 70.4 322 13 BY520186 BY520186  
 191 17.6 70.4 332 10 AW354460 AW354460 36297.MAR  
 192 17.6 70.4 335 29 C8543231 C8543231 CH240-424  
 193 17.6 70.4 360 12 B8755679 B8755679 K-EST0033  
 194 17.6 70.4 381 9 AV649893 AV649893  
 195 17.6 70.4 388 9 AA560373 AA560373 V1390407.r  
 196 17.6 70.4 415 9 AI926861 AI926861 WO60803.x  
 197 17.6 70.4 415 10 B8723040 B8723040 192375.MA  
 198 17.6 70.4 419 28 B31301 B31301 HS-1008-A1-  
 199 17.6 70.4 420 12 B1848567 B1848567 471054.MA  
 200 17.6 70.4 431 10 AW492859 AW492859 UI-M-BH3-  
 201 17.6 70.4 432 13 BY453938 BY453938  
 202 17.6 70.4 459 28 BH391023 BH391023 SP\_0073.A  
 203 17.6 70.4 459 28 A2163159 A2163159 AG-ND-125  
 204 17.6 70.4 465 28 B36475 B36475 HS-1040-B2-  
 205 17.6 70.4 470 14 C8301912 C8301912 TGETSTyGO  
 206 17.6 70.4 478 28 B2866206 B2866206 CH240-388  
 207 17.6 70.4 483 29 BX128188 BX128188 Danio rer  
 208 17.6 70.4 490 10 BF707367 BF707367 282930.MA  
 209 17.6 70.4 500 29 C8762266 C8762266 CH240-2A1  
 210 17.6 70.4 500 9 AA567179 AA567179 AG01227.5  
 211 17.6 70.4 508 9 AA690183 AA690183 vt82802.s  
 212 17.6 70.4 516 12 BM087906 BM087906 500765.MA  
 213 17.6 70.4 525 29 BX203562 BX203562 Danio rer  
 214 17.6 70.4 526 29 BX187137 BX187137  
 215 17.6 70.4 533 10 BF287824 BF287824 EST42415  
 216 17.6 70.4 551 29 BX186681 BX186681 Danio rer  
 217 17.6 70.4 556 10 B8666585 B8666585 150089.MA  
 218 17.6 70.4 562 12 BQ039661 BQ039661 gtd1d09.y  
 219 17.6 70.4 563 29 C8123034 C8123034 tigr-gss-  
 220 17.6 70.4 583 29 BX134737 BX134737 Danio rer  
 221 17.6 70.4 587 29 BX120204 BX120204  
 222 17.6 70.4 572 9 AI666931 AI666931 fc43a04.x

C 224	17.2	68.8	580	9	AV610420	AV610420 AV610420
C 225	17.2	68.8	589	29	CE722891	CE722891 tigr-gss-
C 226	17.2	68.8	597	29	EX137629	EX137629 Dantio rer
C 227	17.2	68.8	602	9	AL957092	AL957092 AL957092
C 228	17.2	68.8	605	12	B1169645	B1169645 RE10883.5
C 229	17.2	68.8	609	12	BJ580769	BJ580769 BJ580769
C 230	17.2	68.8	613	9	AL862644	AL862644 AL862644
C 231	17.2	68.8	619	29	FR002971	Z86754 F.rubripes
C 232	17.2	68.8	629	29	BB252309	BB252309 BB252309
C 233	17.2	68.8	635	29	DR38C85	AL966525 Dantio rer
C 234	17.2	68.8	646	28	AQ310531	AQ310531 CITBI-EI-
C 235	17.2	68.8	651	12	B1228645	B1228645 RE26254.5
C 236	17.2	68.8	651	29	EX229400	EX229400 Dantio rer
C 237	17.2	68.8	654	12	B1485671	B1485671 RE68925.5
C 238	17.2	68.8	655	29	EX153639	EX153639 Dantio rer
C 239	17.2	68.8	658	28	BH381917	BH381917 AG-ND-173
C 240	17.2	68.8	668	28	EX144849	EX144849 Dantio rer
C 241	17.2	68.8	675	29	EX207834	EX207834 Dantio rer
C 242	17.2	68.8	679	12	BG340998	BG340998 602462731
C 243	17.2	68.8	683	29	EX200167	EX200167 Dantio rer
C 244	17.2	68.8	684	29	EX141348	EX141348 Dantio rer
C 245	17.2	68.8	708	29	EX208901	EX208901 Dantio rer
C 246	17.2	68.8	709	29	EX236419	EX236419 Dantio rer
C 247	17.2	68.8	712	13	CA117130	CA117130 SCFRLR100
C 248	17.2	68.8	712	29	CE072314	CE072314 tigr-gss-
C 249	17.2	68.8	717	12	B1860382	B1860382 603385035
C 250	17.2	68.8	727	29	EX223174	EX223174 Dantio rer
C 251	17.2	68.8	743	29	EX210325	EX210325 Dantio rer
C 252	17.2	68.8	744	12	BJ610127	BJ610127 BJ610127
C 253	17.2	68.8	749	28	BZ201364	BZ201364 CH230-445
C 254	17.2	68.8	750	12	BG175998	BG175998 602335223
C 255	17.2	68.8	752	29	CE056270	CE056270 tigr-gss-
C 256	17.2	68.8	767	10	BF238803	BF238803 601304402
C 257	17.2	68.8	785	13	BU396145	BU396145 603805848
C 258	17.2	68.8	789	29	EX192564	EX192564 Dantio rer
C 259	17.2	68.8	793	29	EX178813	EX178813 Dantio rer
C 260	17.2	68.8	798	29	EX181121	EX181121 Dantio rer
C 261	17.2	68.8	804	10	BF303575	BF303575 602031192
C 262	17.2	68.8	808	12	BG281474	BG281474 602401993
C 263	17.2	68.8	809	14	CF147554	CF147554 AGENCOURT
C 264	17.2	68.8	822	12	BM415551	BM415551 OP20627 M
C 265	17.2	68.8	822	14	CA968911	CA968911 CcLLQ3a10
C 266	17.2	68.8	828	12	BI736630	BI736630 603360209
C 267	17.2	68.8	855	13	BU594379	BU594379 AGENCOURT
C 268	17.2	68.8	859	12	BG330583	BG330583 602560171
C 269	17.2	68.8	873	29	CNS021RO	AL199293 Tet-raodon
C 270	17.2	68.8	877	10	BF679992	BF679992 602154705
C 271	17.2	68.8	899	14	CF878053	CF878053 tric016xe
C 272	17.2	68.8	914	13	BU946652	BU946652 703769886
C 273	17.2	68.8	931	10	BE885887	BE885887 601507518
C 274	17.2	68.8	932	29	CNS01U03	AL167880 Tet-raodon
C 275	17.2	68.8	939	14	CA968251	CA968251 CcLM07a27
C 276	17.2	68.8	940	12	BG445132	BG445132 GA_Ea002
C 277	17.2	68.8	948	14	CA968273	CA968273 CcLM07a27
C 278	17.2	68.8	951	14	CA966594	CA966594 CcLM07a28
C 279	17.2	68.8	953	13	BU246550	BU246550 603591487
C 280	17.2	68.8	968	10	BE053652	BE053652 GA_Ea003
C 281	17.2	68.8	971	12	BG253057	BG253057 602365582
C 282	17.2	68.8	980	13	BQ926020	BQ926020 AGENCOURT
C 283	17.2	68.8	982	13	BQ927533	BQ927533 AGENCOURT
C 284	17.2	68.8	989	10	BF793186	BF793186 602253026
C 285	17.2	68.8	993	29	CNS04C2O	AL285117 Tet-raodon
C 286	17.2	68.8	1004	14	CF224572	CF224572 AGENCOURT
C 287	17.2	68.8	1031	28	BZ561787	BZ561787 pac82-164
C 288	17.2	68.8	1042	28	BZ568252	BZ568252 pac82-164
C 289	17.2	68.8	1058	28	CC204139	CC204139 CH261-129
C 290	17.2	68.8	1194	12	BG497858	BG497858 602543153
C 291	17.2	68.8	1259	12	EM55761	EM55761 AGENCOURT
C 292	17.2	68.8	1733	12	BG340667	BG340667 602462246
C 293	17.2	68.8	2090	11	AK043049	AK043049 Mus muscu
C 294	17	68.0	119	12	BG463673	BG463673 EML_50_F0
C 295	17	68.0	131	10	BE063655	BE063655 RC6-BT029
C 296	17	68.0	225	10	BF854792	BF854792 RC6-FN008

297

298

C 299

C 300

17

68.0

234

28

AZ397194

1M0162M01

CE852473 tigr-gss-

BE070638 CM4-BT041

AA317004 EST18869

ALIGNMENTS

BE309187

601095072F1 NCI\_CGAP\_Mam5 Mus musculus cdna clone IMAGE:3489664 5',

mrna sequence.

BE309187

BE309187.1 GI:9166997

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 675)

NIH-MGC http://mgc.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLAM8530 row: p column: 17

High quality sequence stop: 522.

Location/Qualifiers

1..675

/organism="Mus musculus"

/mol\_type="mrna"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:3489664"

/tissue\_type="tumor, gross tissue"

/dev\_stage="7 months"

/lab\_host="DH10B"

/clone\_lib="NCI CGAP Mam5"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

Query Match

Best Local Similarity

Matches

21;

Conservative

0;

Mismatches

3;

Indels

0;

Gaps

0;

76.8%;

Score

19.2;

DB

10;

Length

675;

QY

1

CCCCGAGGCCAGAGGCACATAG 24

|||||

DB

516

CACGAGGCGCCAGAGACCTAG 493

RESULT 2

CF242277/c

LOCUS

CF242277

861 bp

mrna

linear

EST 05-AUG-2003

AGENCOURT 15102430 NICHD\_XGC Emb7 Silurana tropicalis cdna clone

IMAGE:6976306 5', mRNA sequence.

DEFINITION

CF242277

EST.

CF242277.1 GI:33445485

EST.

Silurana tropicalis (western clawed frog)

Silurana tropicalis

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Silurana.

# REFERENCE AUTHORS TITLE JOURNAL COMMENT

1 (bases 1 to 861)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Robert M. Grainger  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM14629 row: e column: 09  
High quality sequence stop: 668.

# FEATURES source

Location/Qualifiers  
1..861  
/organism="Silurana tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="IMAGE:6976306"  
/tissue\_type="tailbud"  
/dev\_stage="embryo, stages 20-27"  
/lab\_hosts="DH10B (phage-resistant)"  
/clone\_lib="NICHD XGC-Emb7"  
/note="Vector: pCMV-SPORT6.1; Site1: NotI; Site2: EcoRV; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb. Constructed by Invitrogen. Note: This is a Xenopus Gene Collection (XGC) library."

# ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 861;  
Best Local Similarity 87.5%; Pred. No. 2.3e+03;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCCGAGGCCAGAGGGCACATAG 24  
DB 861 CACGAGGCCAGTGGGCACAAAG 838

# RESULT 3

BJ036626  
LOCUS BJ036626 452 bp mRNA linear EST 26-SEP-2003  
DEFINITION laevis cDNA clone XL036p06 5', mRNA sequence.  
ACCESSION BJ036626  
VERSION BJ036626.1 GI:17397211  
KEYWORDS EST.  
SOURCE Xenopus laevis (African clawed frog)

# ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.

# REFERENCE AUTHORS

1 (bases 1 to 452)  
Kitayama A., Takasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.

# TITLE JOURNAL COMMENT

Expressed genes in X. laevis embryo  
Unpublished (2001)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: [tsnini@gene.nig.ac.jp](mailto:tsnini@gene.nig.ac.jp)  
The information of this clone is available through the following URL.

# FEATURES source

<http://xenopus.nibb.ac.jp>.  
Location/Qualifiers  
1..452  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL036p06"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 15"  
/clone\_lib="NIBB Mochii normalized Xenopus neurula library"

# ORIGIN

Query Match 75.2%; Score 18.8; DB 12; Length 452;  
Best Local Similarity 90.9%; Pred. No. 2.7e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCCGAGGCCAGAGGGCACAT 22  
DB 345 CTCGAGAGCCAGAGGGCACAT 366

# RESULT 4

AW645104  
LOCUS AW645104 637 bp mRNA linear EST 26-APR-2001  
DEFINITION cm48h12.w1 Blackshear/Soares normalized Xenopus egg library Xenopus laevis cDNA clone PBX0147H12 5', mRNA sequence.

# ACCESSION VERSION KEYWORDS SOURCE ORGANISM

AW645104  
AW645104.1 GI:7402520  
EST.  
Xenopus laevis (African clawed frog)  
Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.

# REFERENCE AUTHORS

1 (bases 1 to 637)  
Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G. Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman, J.W., Bonaldo, M.F. and Soares, M.B.  
The NIEHS Xenopus maternal EST project: interim analysis of the first 13,879 ESTs from unfertilized eggs

# JOURNAL MEDLINE PUBMED COMMENT

Gene 267 (1), 71-87 (2001)  
21211403  
11311557  
Contact: Perry J. Blackshear  
Office of Clinical Research and Laboratory of Signal Transduction  
National Institute of Environmental Health Sciences  
A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709, USA

Tel: 919 541-4899

Fax: 919 541-4571

Email: [black009@niehs.nih.gov](mailto:black009@niehs.nih.gov)

Clone is available through Research Genetics, Inc., 2130 Memorial Parkway, Huntsville, AL 35901  
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email [cdna@resgen.com](mailto:cdna@resgen.com)

DNA Sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC).

PCR Primers:

FORWARD: TGTAAACACGAGCCAGCT

BACKWARD: CAGGAACAGCTATGACC

Plate: 0147 row: H column: 12

Seq primer: T7 primer.

# FEATURES source

Location/Qualifiers  
1..637  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="PBX0147H12"  
/sex="female"  
/tissue\_type="unfertilized egg"  
/cell\_type="unfertilized egg"  
/dev\_stage="unfertilized egg"



/lab\_host="DH10B"  
/clone\_lib="Blackshear/Soares normalized Xenopus egg  
library"  
/note=vector: pT7T3-Pac; Site 1: EcoRI; Site 2: NotI;  
PolyA-selected mRNA was prepared from unfertilized Xenopus  
laevis eggs. The library was constructed in the vector  
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and  
Soares, M.B. 'Normalization and subtraction: two  
approaches to facilitate gene discovery', Genome Research  
6:791-806, 1996. The first strand synthesis used a  
NotI-dn18 primer; double stranded cDNAs were ligated to  
EcoRI adapters, digested with NotI, and directionally  
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.  
The library contained approximately 7.2 X 10<sup>5</sup>  
recombinants, with average insert sizes of 1-1.5 kb."

ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 637;  
Best Local Similarity 90.9%; Pred. No. 3e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CCGGAGGCGCAGAGGCGACAT 22  
Db 360 CTCGAGAGCCAGAGGCGACAT 381

RESULT 5

AZ189837  
LOCUS  
DEFINITION  
SP\_1015 A2\_F12\_SP6E Strongylocentrotus purpuratus, purple sea  
urchin, sperm genomic BAC library Strongylocentrotus purpuratus  
genomic clone Plates=1015 Col=24 Row=K, genomic survey sequence.

ACCESSION

AZ189837  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Strongylocentrotus purpuratus  
Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoida; Euechinoida; Echinacea; Echinoida;  
Strongylocentrotidae; Strongylocentrotus.

REFERENCE

Ameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,  
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T.,  
Wray, G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H.  
and Hood, L.  
A sea urchin genome project: Sequence scan, virtual map, and  
additional resources

TITLE

Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)  
20402566  
10920195

JOURNAL

Medline  
PUBLISHED  
COMMENT  
Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 1015 row: K column: 24  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 760.

FEATURES

source  
Location/Qualifiers  
1..760  
/organism="Strongylocentrotus purpuratus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7668"  
/clone\_lib="Plates=1015 Col=24 Row=K"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea  
urchin, sperm genomic BAC library"  
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli  
DH10B"

ORIGIN

Query Match 75.2%; Score 18.8; DB 28; Length 760;  
Best Local Similarity 90.9%; Pred. No. 3.2e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 CCGAGGCGCAGAGGCGACATAG 24  
Db 600 CCGAGGCGCAGAGGCGACATAG 621

RESULT 6

BH312135  
LOCUS  
DEFINITION  
CH230-101J1.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
CH230-101J1, genomic survey sequence.

ACCESSION

BH312135  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE

1 (bases 1 to 768)  
Zhao, S., Shetty, J., Shatsman, S., Tsengaye, G., Geer, K.,  
Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,  
Riggs, F., de Jong, P. and Fraser, C.M.  
Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
Unpublished (1999)

TITLE

Other\_GSSs: CH230-101J1.TJ  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org

JOURNAL

Clones are derived from the rat BAC library CHORI-230  
(http://www.chori.org/bacpac/rat230.htm). For BAC library  
availability, please contact Pieter de Jong (pdejong@mail.cho.org).  
Clones may be purchased from BACPAC Resources  
(http://www.chori.org/bacpac/or ering information.htm). BAC end  
page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
Plate: 101 row: J column: 1  
Seq primer: T7  
Class: BAC ends.

FEATURES

source  
Location/Qualifiers  
1..768  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SSNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-101J1"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_lib="CHORI-230 Segment 1"  
/note="Vector: PTAB3AC2.1; Site 1: EcoRI; Site 2: EcoRI;  
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
Pieter de Jong"

ORIGIN

Query Match 75.2%; Score 18.8; DB 28; Length 768;  
Best Local Similarity 90.9%; Pred. No. 3.2e+03;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 CCGGAGGCGCAGAGGCGACAT 22  
Db 745 CCGGAGGCGCAGAGGCGACATAT 766

RESULT 7

BF248203  
LOCUS  
DEFINITION  
BF248203  
601821202F1 NIH\_MGC\_62 Homo sapiens cDNA clone IMAGE:4046136 5',  
linear EST 14-NOV-2000

```

mRNA sequence.
ACCESSION BF248203
VERSION BF248203.1 GI:11164466
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL 1 (bases 1 to 866)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM873 row: c column: 01
High quality sequence stop: 402.
FEATURES
source
1..866
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4046136"
/tissue_type="melanotic melanoma, high MDR"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_62"
/note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgccctggcc); Site_2: SfiI (ggccattatggcc);
Double-stranded cDNA was prepared from cell line RNA.
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGCGGCGGCATG-DT(30)BN-3',
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
ORIGIN
Query Match 75.2%; Score 18.8; DB 10; Length 866;
Best Local Similarity 90.9%; Pred.No. 3.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 3 CGGAGGGCCAGGCGGCACATAG 24
Db 523 CGGAGGGCCAGGCGGCACAG 544
RESULT 8
CB587782
LOCUS CB587782
DEFINITION AGENCOURT 12974016 NIH_MGC 136 Mus musculus cDNA clone
IMAGE:30293489 5', mRNA sequence.
ACCESSION CB587782
VERSION CB587782.1 GI:29505638
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL 1 (bases 1 to 866)
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. David Rowe
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDAM339 row: a column: 18
High quality sequence start: 16
High quality sequence stop: 485.
FEATURES
source
1..866
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30293489"
/tissue_type="embryonic limb, maxilla and mandible"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_136"
/note="Vector: PCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
Normalized, full-length enriched library from pool of
mouse embryonic limb, maxilla and mandible, embryonic day
17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
equivalents from respective days). Cloned directionally,
oligo-dT primed (5'-GACTAGTCTAGATCGGAGCGGCGGC(T)15-3'.
Size selected for the >1kb fragments, average insert size
1.2 kb. Normalization to Cot 7.5. Tissue contributed by
David Rowe; library constructed by Resgen, Invitrogen
Corp. Note: this is a NIH_MGC Library."
ORIGIN
Query Match 75.2%; Score 18.8; DB 14; Length 866;
Best Local Similarity 90.9%; Pred.No. 3.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CCCGAGGGCCAGGCGGCACAT 22
Db 664 CCCTGAGGGCCAGGCGGCACAT 685
RESULT 9
CD215458/c
LOCUS CD215458
DEFINITION pppln.pk002.g9 Normalized chicken pituitary/hypothalamus/pineal
cDNA library (pppln) Gallus gallus cDNA clone pppln.pk002.g9 5',
similar to db|BAA92562.1 (AB037745) KIAA1324 protein [Homo
sapiens], mRNA sequence.
ACCESSION CD215458
VERSION CD215458.1 GI:30955457
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Porter, T.E. and Cogburn, L.A.
TITLE Chicken ESTs from pituitary/hypothalamus/pineal gland
JOURNAL Unpublished (2001)
COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES
source
1..236
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="pppln.pk002.g9"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"

```

prepare full-length cDNA libraries  
Normalization and subtraction  
genes. Genome Res. 10 (10), 161  
RKEN integrated sequence anal-  
sequencing pipeline with 384 mul-  
10 (11) 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001).  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers  
 1. .326  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="L630023J01"  
 /sex="male"  
 /tissue\_type="brain"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="RIKEN full-length enriched, adult male brain"  
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATCCGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using thermostable thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCGAGTAAATTAATCCCGCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC 1"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 326;  
 Best Local Similarity 84.0%; Pred. No. 2.9e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGGCACATAGC 25

DB 57 CCGGAGGGCCGATAGGCACAGAGC 33

## RESULT 12

BY008855/c

LOCUS

DEFINITION BY008855 RIKEN full-length enriched, 16 days neonate male diencephalon Mus musculus cDNA clone G63003K24 5', mRNA sequence.

ACCESSION

BY008855

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 351)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yananaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Buit, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Bruscia, V., Chotha, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasaki, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,

## TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome.res@gs.riken.go.jp  
 URL: <http://genome.gsc.riken.go.jp/>  
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numasaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submision  
 Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

## FEATURES

source

Location/Qualifiers

1. .351  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="G63003K24"  
 /sex="male"  
 /tissue\_type="diencephalon"  
 /dev\_stage="16 days neonate"  
 /clone\_lib="RIKEN full-length enriched, 16 days neonate male diencephalon"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 351;  
 Best Local Similarity 84.0%; Pred. No. 3e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY

1 CCGGAGGGCCAGAGGGCACATAGC 25

Db	61	CCGGAGGGCCGCTAGGCACAGAGC	37	Best Local Similarity 84.0%; Pred. No. 3e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
RESULT 13	BB800951/c	BB800951	RIKEN full-length enriched, 15 days neonate male	EST 16-NOV-2001
LOCUS	BB800951	BB800951	diencephalon Mus musculus cDNA clone G630033K24 3', mRNA sequence.	
DEFINITION	BB800951	BB800951	diencephalon Mus musculus cDNA clone G630033K24 3', mRNA sequence.	
ACCESSION	BB800951	BB800951	BB800951	BB800951
VERSION	BB800951.1	BB800951.1	GI:16970574	GI:16970574
KEYWORDS	EST.	EST.	EST.	EST.
SOURCE	Mus musculus	Mus musculus	(house mouse)	
ORGANISM	Mus musculus	Mus musculus		
REFERENCE	1	(Bases 1 to 358)		
AUTHORS	Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiraoaka, K., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyata, T., Watahiki, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.			
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (AKIMURA, T., et al. 2001)			
JOURNAL	Unpublished (2001)			
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/ Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10): 1617-1630 (2000) wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11): 1757-1771 (2000) Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2): 281-289 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues.			
FEATURES	Location/Qualifiers	1..358		
source	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/strain="C57BL/6J"			
	/db_xref="taxon:10090"			
	/clone="G630033K24"			
	/sex="male"			
	/tissue type="diencephalon"			
	/dev stages="16 days neonate"			
	/clone_libs="RIKEN full-length enriched, 16 days neonate male diencephalon"			
ORIGIN	Query Match 74.4%; Score 18.6; DB 10; Length 358;			
	Best Local Similarity 84.0%; Pred. No. 3.5e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY	1	CCGGAGGGCCGCTAGGCACAGAGC	37	
Db	61	CCGGAGGGCCGCTAGGCACAGAGC	37	
RESULT 14	BM540159	BM540159	linear	EST 20-FEB-2002
LOCUS	BM540159	BM540159	hbl8g12.g1 Canis cDNAs from testes cells Canis familiaris cDNA	
DEFINITION	BM540159	BM540159	hbl8g12.g1 Canis cDNAs from testes cells Canis familiaris cDNA	
ACCESSION	BM540159	BM540159	BM540159	BM540159
VERSION	BM540159.1	BM540159.1	GI:18822713	GI:18822713
KEYWORDS	EST.	EST.	EST.	EST.
SOURCE	Canis familiaris	Canis familiaris	(dog)	
ORGANISM	Canis familiaris	Canis familiaris		
REFERENCE	1	(Bases 1 to 584)		
AUTHORS	O'Shaughnessy, A.L., McCombie, W.R., Baker, J.P., Balija, V., Cunniss, D., Dedhia, N.N., de la Bastide, M., Katzenberger, F., King, L., Kirchoff, K.A., Miller, B., Muller, S., Nascimento, L.U., Palmer, L., Santos, L., Shah, R.S., Spiegel, L.A., Zutavern, I., Preston, R., and Hannon, G.J.			
TITLE	Expressed sequence tags from Canis familiaris (dog) (2002)			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: W. Richard McCombie Lita Annenberg Hazen Genome Sequencing Center Cold Spring Harbor Laboratory PO Box 100, Cold Spring Harbor, NY 11724, USA Tel: 516 367 8884 Fax: 516 367 8874 Email: mcombie@cshl.org Plate: hb18 row: g column: 12 Seq primer: -21M13UnivRev High quality sequence stop: 584. Location/Qualifiers			
FEATURES	Location/Qualifiers	1..584		
source	/organism="Canis familiaris"			
	/mol_type="mRNA"			
	/db_xref="taxon:9615"			
	/clone="hbl8g12"			
	/tissue type="testes"			
	/clone_libs="Canis cDNAs from testes cells"			
	/note="Vector: Lambda Zap II; The library was produced by Greg Hannon and Raymond Preston (Cold Spring Harbor Laboratory). This library is oligo(dT) primed using stragene zap cDNA synthesis kit. It was made from dog testes. Please contact Greg Hannon (hannon@cshl.org) with any library related inquiries."			
ORIGIN	Query Match 74.4%; Score 18.6; DB 12; Length 584;			
	Best Local Similarity 84.0%; Pred. No. 3.5e+03; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY	1	CCCGAGGGCCGAGGGCACATAGC	25	
Db	122	CCCGAGGGCCGAGGGCACATAGC	146	
RESULT 15	BM538550	BM538550	625 bp	EST 20-FEB-2002
LOCUS	BM538550	BM538550	ha97a10.g1 Canis cDNAs from testes cells Canis familiaris cDNA	
DEFINITION	BM538550	BM538550	clone ha97a10 5', mRNA sequence.	
ACCESSION	BM538550	BM538550	BM538550	BM538550
VERSION	BM538550.1	BM538550.1	GI:18819724	GI:18819724
KEYWORDS	EST.	EST.	EST.	EST.
SOURCE	Canis familiaris	Canis familiaris	(dog)	
ORGANISM	Canis familiaris	Canis familiaris		

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

#### REFERENCE AUTHORS

O'Shaughnessy,A.L., McCombie,W.R., Baker,J.P., Balija,V.,  
Cunning,D., Dedhia,N.N., de la Bastide,M., Katzenberger,F.,  
King,L., Kirchoff,K.A., Miller,B., Muller,S., Nascimento,L.U.,  
Palmer,L., Santos,L., Shah,R.S., Spiegel,L.A., Zutavern,T.,  
Preston,R. and Hamon,G.J.  
Expressed sequence tags from Canis familiaris (dog) (2002)

#### TITLE JOURNAL COMMENT

Unpublished (2002)  
Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100 Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874

Email: mcombie@cshl.org  
Plate: ha97 row: a column: 10  
Seq primer: -21M3UnivRev  
High quality sequence stop: 625.

#### FEATURES

source

1..625  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/db\_xref="taxon:9615"  
/clone="ha97a10"  
/tissue\_type="testes"  
/clone\_lib="Canis cDNAs from testes cells"  
/note="Vector: Lambda Zap II; The library was produced by  
Greg Hannon and Raymond Preston (Cold Spring Harbor  
Laboratory). This library is oligo(dT) primed using  
stratagene zap cDNA synthesis kit. It was made from dog  
testes. Please contact Greg Hannon (hannon@cshl.org) with  
any library related inquiries."

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 12; Length 625;  
Best Local Similarity 84.0%; Pred. No. 3.6e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGCGCCAGAGGCACATAGC 25

Db 261 CCGGATGACCGAGTGGCAGATAGC 285

#### RESULT 16

BB657042/c

LOCUS BB657042 RIKEN full-length enriched, 12 days embryo spinal ganglion  
DEFINITION Mus musculus cDNA clone D130072G11 5', mRNA sequence.

ACCESSION BB657042

VERSION BB657042.1 GI:16490868

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

#### REFERENCE AUTHORS

1 (bases 1 to 637)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,  
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,  
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,  
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,  
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
Takeda,Y., Tanaka,T., Taya,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T. et al. 2001)

#### TITLE JOURNAL COMMENT

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222

Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,  
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.  
and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,  
Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,  
Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

#### FEATURES source

Location/Qualifiers  
1..637  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="D130072G11"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"

/note="Site 1: Sall; Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGCGCGCGCAACTCGAGTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGAGATCTCGAGTCTTAATAATATCCCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

#### ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 637;  
Best Local Similarity 84.0%; Pred. No. 3.6e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGCGCCAGAGGCACATAGC 25

Db 31 CCGGAGGCGCGCTAGGGCAGAGGC 7

#### RESULT 17

BI652362/c

LOCUS BI652362

DEFINITION

603299633F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5340176 5',

RNA sequence.

ACCESSION BI652362

VERSION BI652362.1 GI:15566598

KEYWORDS EST.

BI652362 651 bp mRNA linear EST 12-SEP-2001





genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.

# FEATURES

source  
Location/Qualifiers

1..734  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="168E09"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG168AC05LP1-end : T7"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 29; Length 734;  
Best Local Similarity 84.0%; Pred. No. 3.8e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGCCCGAGGGCAGCATAGC 25  
Db 454 CCCGTCGGCGAGAGCAGCATAGC 478

## RESULT 20

LOCUS BI109078/c 825 bp mRNA linear EST 26-JUN-2001  
DEFINITION 602896924F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:5027164 5',  
mRNA sequence.

ACCESSION BI109078.1 GI:14559979

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 825)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Prepared by: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1076 row: 0 column: 05

High quality sequence start: 11

High quality sequence stop: 767.

## FEATURES

source

1..825  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5027164"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam5"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 12; Length 825;  
Best Local Similarity 84.0%; Pred. No. 4e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGCCCGAGGGCAGCATAGC 25  
Db 454 CCCGTCGGCGAGAGCAGCATAGC 478

Db 36 CCGGAGGCCCGTAGGGCAGCAGC 12

## RESULT 21

LOCUS BI646763/c

DEFINITION 603279793F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5320031 5',  
mRNA sequence.

ACCESSION BI646763.1 GI:15560999

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 839)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
cDNA Library Prepared by: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM1811 row: 1 column: 24

High quality sequence stop: 823.

## FEATURES

source

1..839  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="129, C57BL/6J, FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5320031"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="10 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP Mam3"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH  
Reference for transgenic model: Xu et al., Nature Genetics  
22, 37-43 (1999)."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 12; Length 839;

Best Local Similarity 84.0%; Pred. No. 4e+03;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGCCCGAGGGCAGCATAGC 25

Db 820 CCCAAAGTCCAGAGGGCAGCAGC 796

## RESULT 22

LOCUS CNS02THM

DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
164107 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION AL213187

VERSION AL213187.1 GI:7872006

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Tetraodon.



```

REFERENCE
AUTHORS
1 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Eizanes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL
MEDLINE
PUBMED
20296633
10835645
REFERENCE
AUTHORS
2 Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
JOURNAL
MEDLINE
PUBMED
20359837
10899143
REFERENCE
AUTHORS
3 (bases 1 to 945)
Genoscope.
TITLE
Direct Submission
JOURNAL
MEDLINE
PUBMED
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Source
Location/Qualifiers
1..945
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone_lib="G"
/clone_lib="G"
/note="Genoscope sequence ID : COAG1644RF04LP1-end : T7"
ORIGIN
Query Match 74.4%; Score 18.6; DB 29; Length 945;
Best Local Similarity 84.0%; Pred. No. 4.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGAGGGCCAGAGGCGCATAGC 25
Db 692 CCGGTCGGGAGAGAGCATAGC 716
RESULT 23
BE887139
LOCUS
DEFINITION
601508337F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3909969 5',
mRNA sequence.
ACCESSION
BE887139
VERSION
BE887139.1 GI:10342129
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 1050)
TITLE
NIH-MGC http://img.nci.nih.gov/.
JOURNAL
MEDLINE
PUBMED
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabs@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9724 row: 1 column: 10
FEATURES
Source
Location/Qualifiers
1..1050
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3909969"
/tissue_type="leiomyosarcoma"
/lab_host="pH108 (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: Not1;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 1050;
Best Local Similarity 84.0%; Pred. No. 4.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CCGGAGGGCCAGAGGCGCATAGC 25
Db 721 CCGGAGGGCCCTGAGGTCCTATGCC 745
RESULT 24
AK051753/c
LOCUS
DEFINITION
Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length
enriched library, clone: D130072G11 product: histone deacetylase 2,
full insert sequence.
ACCESSION
AK051753
VERSION
AK051753.1 GI:26342185
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
1 Carrincci,P. and Hayashizaki,Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
MEDLINE
PUBMED
99279253
10349636
REFERENCE
AUTHORS
2 Carrincci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL
MEDLINE
PUBMED
20499374
11042159
REFERENCE
AUTHORS
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,I., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL
MEDLINE
PUBMED
20530913
11076861
REFERENCE
AUTHORS
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
MEDLINE
PUBMED
Nature 409, 685-690 (2001)
REFERENCE
AUTHORS
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation

```

JOURNAL REFERENCE AUTHORS	of 60,770 full-length cDNAs		Accession	
	Nature 420, 563-573 (2002)		BZ344555.1 GI:24901698	
	6 (bases 1 to 3251)		GSS	
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imorani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.		Sorghum bicolor (sorghum)	
TITLE JOURNAL	Direct Submission		Sorghum bicolor (sorghum)	
	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome		Sorghum bicolor (sorghum)	
	Exploitation Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)		Sorghum bicolor (sorghum)	
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.		Sorghum bicolor (sorghum)	
COMMENT	Please visit our web site for further details.		Sorghum bicolor (sorghum)	
	URL: http://genome.gsc.riken.go.jp/		Sorghum bicolor (sorghum)	
	URL: http://fantom.gsc.riken.go.jp/		Sorghum bicolor (sorghum)	
	Location/Qualifiers		Sorghum bicolor (sorghum)	
FEATURES	1..3251		Sorghum bicolor (sorghum)	
	/organism="Mus musculus"		Sorghum bicolor (sorghum)	
	/mol_type="RNA"		Sorghum bicolor (sorghum)	
	/strain="C57BL/6J"		Sorghum bicolor (sorghum)	
CDS	/db_xref="FANTOM DB:D130072G11"		Sorghum bicolor (sorghum)	
	/db_xref="MG1:2420128"		Sorghum bicolor (sorghum)	
	/db_xref="taxon:10090"		Sorghum bicolor (sorghum)	
	/clone="D130072G11"		Sorghum bicolor (sorghum)	
ORIGIN	/tissue_type="spinal ganglion"		Sorghum bicolor (sorghum)	
	/clone_lib="RIKEN full-length enriched mouse cDNA library"		Sorghum bicolor (sorghum)	
	/dev_stage="12 days embryo"		Sorghum bicolor (sorghum)	
	184..1098		Sorghum bicolor (sorghum)	
JOURNAL REFERENCE AUTHORS	/note="unnamed protein product; histone deacetylase 2 (MGD MG1:1097691, GB NM_008229, evidence: BLASTN, 99%, match=1011) putative"		Sorghum bicolor (sorghum)	
	/codon_start=1		Sorghum bicolor (sorghum)	
	/protein_id="BAC34755.1"		Sorghum bicolor (sorghum)	
	/db_xref="GI:26342186"		Sorghum bicolor (sorghum)	
TITLE JOURNAL	/translation="MAYSGGGKKKVCYVYDGDIGNYYGGHPMKPHRIMTHNLLL NYGLYRKMEIYRPHGTAEEMTKYSDIEIRPLRSIRPDNMSEYKQMRPNVEDCP VFDGFEFCOLYIDIDTHGVDGEAEPTTDRTVMTVSFKYGEYFPGTGLRIDGAG KKKYAVNFPWPDGDDSDYSGQLFKPIISKWMYQPSAVVLQCGADSLSDGLGCFN LTVKGNQTFPLVWCFSFPRSHKSPHC"		Sorghum bicolor (sorghum)	
	74.4%; Score 18.6; DB 11; Length 3251;		Sorghum bicolor (sorghum)	
	Best Local Similarity 84.0%; Pred. No. 6.3e+03;		Sorghum bicolor (sorghum)	
	Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;		Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
RESULT 25 BZ344555 LOCUS	350 bp DNA linear GSS 12-NOV-2002		Sorghum bicolor (sorghum)	
	hq87c01.b1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum		Sorghum bicolor (sorghum)	
	bicolor genomic clone hq87c01 5', genomic survey sequence.		Sorghum bicolor (sorghum)	
	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
DEFINITION			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
QY	1 CCCGAGGCCGACAGGGCCATAGC 25		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	
			Sorghum bicolor (sorghum)	
DB	35 CCCGAGGCCGACAGGGCCATAGC 11		Sorghum bicolor (sorghum)	

The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

FEATURES  
Location/Qualifiers  
1..587

/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

## ORIGIN

Query Match 73.6%; Score 18.4; DB 29; Length 587;  
Best Local Similarity 95.0%; Pred. No. 4.3e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGGCACA 21

DB 254 CCTGAGGCCAGAGGGGCACA 273

RESULT 27  
LOCUS  
AL565528/c  
DEFINITION  
AL565528 Homo sapiens FETAL BRAIN Homo sapiens cdna clone  
CSODF004XJ16 3-PRIME, mRNA sequence.

ACCESSION  
AL565528

VERSION  
AL565528.2 GI:30549632

KEYWORDS  
EST.

SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens

REFERENCE  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE  
Full-length cdna libraries and normalization

JOURNAL  
Unpublished (2001)

COMMENT  
On Feb 16, 2001 this sequence version replaced gi:12916994.

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2308.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CSODF004DE08NPI&cluster=2308.f. Contact :

Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CSODF004DE08NPI.

Location/Qualifiers

1..1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CSODF004XJ16"

/tissue\_type="FETAL BRAIN"

/dev\_stage="fetal"

/clone\_lib="Homo sapiens FETAL BRAIN"

/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cdna

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-strand cdna was digested with NotI and

cloned into the NotI and EcoRV sites of the pCMVSPORT 6

vector. Library was not normalized."

## ORIGIN

Query Match 73.6%; Score 18.4; DB 9; Length 1201;

Best Local Similarity 58.3%; Pred. No. 5.4e+03;  
Matches 14; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCGAGGGCCAGAGGGGCACATAG 24

DB 62 SCNVRASSCCSRGAGGGGACATGG 39

## RESULT 28

BB483005/c

LOCUS

DEFINITION

BB483005 RIKEN full-length enriched, 13 days embryo lung Mus

musculus cdna clone D430018A20 3' similar to X06340 Mouse mrna for

p-cadherin, mRNA sequence.

ACCESSION

BB483005

VERSION

BB483005.1 GI:9400614

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 275)

AUTHORS

Kanno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,

Carninci, P., Endo, T., Fukuda, S., Fukuishi, Y., Hara, A., Hayatsu, N.,

Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,

Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H.,

Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K.,

Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,

Takahashi, F., Tominaga, N., Tova, T., Tsunoda, Y., Watabiki, A.,

Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A.,

Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Yuramatsu, M. and

Hayashizaki, Y.

RIKEN Mouse ESTs (Kanno, H., et al.)

Unpublished (2000)

CONTACT: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,

Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermostable enzymes by

trehalose and its application for the synthesis of full length

cdna. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh, M., Kitsumai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,

Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cdna cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

further details.

FEATURES  
Location/Qualifiers  
1..275

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="D430018A20"

/tissue\_type="lung"

/dev\_stage="13 days embryo"

/lab\_host="DH10B"

/clone\_lib="RIKEN full-length enriched, 13 days embryo

lung"

/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5']  
 GAGAGAGAGCGCCGACCTCGAGTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5']  
 GAGAGAGATTCGAGTTAAATTAATATCCCTCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

## ORIGIN

Query Match 72.8%; Score 18.2; DB 10; Length 275;  
 Best Local Similarity 87.0%; Pred. No. 4e+03; 3; Indels 0; Gaps 0;  
 Matches 20; Conservative 0; Mismatches 3

QY 2 CCGGAGGCGCCAGAGGCGACATAG 24

Db 53 CTGGAGGCGCCAGAGGCGAGATAG 31

## RESULT 29

## BY556125/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

BY556125 446 bp mRNA linear EST 14-DEC-2002  
 BY556125 RIKEN full-length enriched, B6-derived CD11 +ve dendritic cells Mus musculus cDNA clone F730316D02 3', mRNA sequence.  
 BY556125  
 BY556125.1 GI:26890504  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 446)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamakata, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Sempke, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verrano, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 22354683  
 12466851  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

## FEATURES

## source

1. .446  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="F730316D02"  
 /cell\_type="B6-derived CD11 +ve dendritic cells"  
 /clone\_lib="RIKEN full-length enriched, B6-derived CD11 +ve dendritic cells"

## ORIGIN

Query Match 72.8%; Score 18.2; DB 13; Length 446;  
 Best Local Similarity 87.0%; Pred. No. 4.7e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGAGGCGCCAGAGGCGACATA 23

Db 249 CCGGAGGCGCCAGAGGCGACATA 227

## RESULT 30

## BY556125/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

BY556125 466 bp mRNA linear EST 12-JUL-2001  
 RE41056.5prime RE Drosophila melanogaster normalized Embryo pFLC-1  
 Drosophila melanogaster cDNA clone RE41056 5 similar to cdt:  
 Pfam003057 'transporter' located on: 2L 23B1-23B1.1: 05/12/2001,  
 mRNA sequence.  
 BY556125  
 BY556125.1 GI:14712301  
 EST.  
 Drosophila melanogaster (fruit fly)  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 466)  
 Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,  
 Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,

The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
 Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H.,  
 Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
 Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
 Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and  
 Hayashizaki, Y. Direct Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics  
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome  
 Trust/MRC Building Addenbrookes Hospital Cambridge) whose  
 assistance we gratefully acknowledge.  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.

George.R., Gonzalez.M., Guarin,H., Harris,N., Li.P., Liao,G.,  
 Misra,S., Mungall,C.J., Nuncio,J., Pacleb,J., Paragas,V., Park,S.,  
 Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and  
 Rubin,G.M.  
 BDGP/HMI RE Drosophila EST Project  
 Unpublished (2001)  
 Contact: Stapleton, M.  
 BDGP

TITLE  
 JOURNAL  
 COMMENT

Lawrence Berkeley National Lab  
 One Cyclotron Rd, Berkeley, CA 94720, USA  
 Fax: 510 486 6798  
 Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)  
 hit genomic AE03582: arm:2L [2452680,2747991]  
 estimated-cyto:22F3-23B3: 05/12/2001  
 Plate: RE.410 row: E column: 8  
 High quality sequence stop: 409.

FEATURES

source

Location/Qualifiers  
 1..466  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7227"  
 /clone="RE41056"  
 /sex="male and female"  
 /dev\_stage="0-24 hours mixed stage embryonic"  
 /lab\_host="DHS-alpha Tona"  
 /clone\_lib="RE Drosophila melanogaster normalized Embryo  
 pFlc-1"  
 /note="Organ: embryo; Vector: pFlc1; Site:1: XhoI; Site:2:  
 BamHI; Library was kindly generated by Piero Carninci at  
 the RIKEN. The library was normalized and excised using  
 Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 72.8%; Score 18.2; DB 12; Length 466;  
 Best Local Similarity 87.0%; Pred. No. 4.7e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CGGAGGGCCAGAGGGCACATAGC 25  
 |||||  
 DB 414 CGGATGGCCAGAGCACATTGC 392

Search completed: June 20, 2004, 14:15:35  
 Job time : 1194.25 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 27.6455 Seconds  
(without alignments)  
501.846 Million cell updates/sec

Title: US-10-624-714-17

Perfect score: 25

Sequence: 1 cccgagggccagagccatagc 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	4403765	3	US-09-103-840A-2
2	25	100.0	4411529	3	US-09-103-840A-1
3	17.2	68.8	16595	4	US-09-146-053-7
4	16.8	67.2	481	4	US-09-621-976-167
5	16.8	67.2	2059	4	US-09-604-978-5
6	16.8	67.2	2059	4	US-09-604-728-5
7	16.6	66.4	533	1	US-08-049-283A-1
8	16.6	66.4	10684	3	US-08-618-100B-3
9	16.4	65.6	1494	4	US-09-277-262-3
10	16.4	65.6	2240	4	US-09-277-262-7
11	16.2	64.8	1262	4	US-09-482-273-44
12	16.2	64.8	1357	3	US-09-448-176-9
13	16.2	64.8	1357	3	US-09-448-176-10
14	16.2	64.8	1357	3	US-09-448-176-11
15	16.2	64.8	3867	3	US-08-762-428A-5
16	16.2	64.8	4016	3	US-08-762-428A-7
17	16	64.0	1278	3	US-09-050-159-123
18	16	64.0	1973	4	US-09-976-594-193
19	16	64.0	2201	4	US-09-566-921-52
20	16	64.0	3600	4	US-09-462-917A-106
21	16	64.0	17606	3	US-08-943-731-4
22	15.8	63.2	1231	4	US-09-015-188-5
23	15.8	63.2	7680	4	US-08-210-748A-3
24	15.8	63.2	49377	1	US-08-764-233A-1
25	15.6	62.4	539	4	US-09-621-976-1421
26	15.6	62.4	1518	4	US-09-719-919A-2
27	15.6	62.4	1613	4	US-09-061-154-1

c	28	15.6	62.4	1613	4	US-09-061-154-3	Sequence 3, Appli
	29	15.6	62.4	1654	3	US-08-931-426-3	Sequence 3, Appli
	30	15.6	62.4	1654	3	US-09-143-470-3	Sequence 3, Appli
	31	15.6	62.4	1654	4	US-08-649-006A-3	Sequence 3, Appli
	32	15.6	62.4	1654	4	US-09-771-023-3	Sequence 2, Appli
	33	15.6	62.4	2153	2	US-08-864-804-2	Sequence 3, Appli
	34	15.6	62.4	2254	4	US-09-138-277C-2	Sequence 2, Appli
	35	15.6	62.4	2447	1	US-08-050-132A-1	Sequence 1, Appli
	36	15.6	62.4	2447	3	US-08-750-222A-1	Sequence 1, Appli
	37	15.6	62.4	2447	3	US-08-815-652B-1	Sequence 1, Appli
	38	15.6	62.4	2447	3	US-08-254-353A-1	Sequence 1, Appli
	39	15.6	62.4	2447	5	PCT-US92-05374A-1	Sequence 1, Appli
	40	15.6	62.4	2447	5	PCT-US95-07084-1	Sequence 1, Appli
	41	15.6	62.4	3133	1	US-08-162-809-1	Sequence 1, Appli
	42	15.6	62.4	4079	4	US-09-016-434-1173	Sequence 1173, Ap
	43	15.6	62.4	4426	4	US-09-658-687A-3	Sequence 3, Appli
	44	15.6	62.4	4546	4	US-09-146-053-6	Sequence 6, Appli
	45	15.4	61.6	333	4	US-09-489-039A-3453	Sequence 3453, Ap
	46	15.4	61.6	345	4	US-09-127-946-5	Sequence 5, Appli
	47	15.4	61.6	345	4	US-09-595-549-11	Sequence 11, Appli
	48	15.4	61.6	531	4	US-09-976-594-436	Sequence 436, Ap
	49	15.4	61.6	555	4	US-09-252-991A-5563	Sequence 5563, Ap
	50	15.4	61.6	807	4	US-09-833-381-42	Sequence 42, Appl
	51	15.4	61.6	1098	4	US-09-170-496D-117	Sequence 117, Appl
	52	15.4	61.6	1098	4	US-09-170-496D-225	Sequence 225, Appl
	53	15.4	61.6	1280	3	US-09-199-737-1	Sequence 1, Appli
	54	15.4	61.6	1280	4	US-08-900-230-1	Sequence 1, Appli
	55	15.4	61.6	1280	4	US-09-058-333A-1	Sequence 1, Appli
	56	15.4	61.6	1299	4	US-09-252-991A-5631	Sequence 5631, Ap
	57	15.4	61.6	1326	4	US-09-252-991A-15735	Sequence 15735, A
	58	15.4	61.6	1330	3	US-08-888-077A-26	Sequence 26, Appl
	59	15.4	61.6	1330	4	US-09-665-479A-15	Sequence 15, Appli
	60	15.4	61.6	1515	4	US-09-252-991A-15701	Sequence 15701, A
	61	15.4	61.6	1597	2	US-08-724-974A-1	Sequence 1, Appli
	62	15.4	61.6	1611	4	US-09-252-991A-5600	Sequence 5600, Ap
	63	15.4	61.6	1697	4	US-09-364-425B-26	Sequence 26, Appl
	64	15.4	61.6	1698	4	US-09-489-039A-1732	Sequence 1732, Ap
	65	15.4	61.6	1794	3	US-08-920-270-1	Sequence 1, Appli
	66	15.4	61.6	1866	3	US-08-908-742-1	Sequence 1, Appli
	67	15.4	61.6	1866	3	US-09-412-289-1	Sequence 1, Appli
	68	15.4	61.6	1981	4	US-09-620-312D-870	Sequence 870, App
	69	15.4	61.6	2118	4	US-09-252-991A-5605	Sequence 5605, Ap
	70	15.4	61.6	2614	2	US-08-795-868-15	Sequence 15, Appl
	71	15.4	61.6	2614	4	US-09-303-089-15	Sequence 15, Appl
	72	15.4	61.6	2614	4	US-09-134-250-15	Sequence 15, Appl
	73	15.4	61.6	2874	4	US-09-489-039A-1500	Sequence 1500, Ap
	74	15.4	61.6	2898	4	US-09-051-696-7	Sequence 7, Appli
	75	15.4	61.6	3901	1	US-08-326-297-1	Sequence 1, Appli
	76	15.4	61.6	3901	3	US-08-617-454-1	Sequence 1, Appli
	77	15.4	61.6	3901	5	PCT-US94-01144-1	Sequence 1, Appli
	78	15.4	61.6	4616	1	US-08-340-203A-1	Sequence 1, Appli
	79	15.4	61.6	4616	2	US-08-452-567-1	Sequence 1, Appli
	80	15.4	61.6	4616	2	US-08-452-567-1	Sequence 1, Appli
	81	15.4	61.6	4616	3	US-09-085-407-1	Sequence 1, Appli
	82	15.4	61.6	6420	2	US-08-374-483-1	Sequence 1, Appli
	83	15.4	61.6	6803	3	US-08-665-259-19	Sequence 19, Appl
	84	15.4	61.6	6803	3	US-08-762-500-19	Sequence 19, Appl
	85	15.4	61.6	8236	1	US-08-461-837-1	Sequence 1, Appli
	86	15.4	61.6	8236	3	US-08-973-223-1	Sequence 1, Appli
	87	15.4	61.6	8236	3	US-09-347-060-1	Sequence 1, Appli
	88	15.4	61.6	8236	5	PCT-US96-09495-1	Sequence 1, Appli
	89	15.4	61.6	9592	1	US-08-393-734-3	Sequence 3, Appli
	90	15.4	61.6	9592	3	US-08-894-489-3	Sequence 3, Appli
	91	15.4	61.6	9641	2	US-08-374-483-3	Sequence 1, Appli
	92	15.4	61.6	10398	2	US-08-331-384-1	Sequence 1, Appli
	93	15.4	61.6	10398	2	US-08-708-188-1	Sequence 1, Appli
	94	15.4	61.6	10398	2	US-08-836-087-1	Sequence 1, Appli
	95	15.4	61.6	10398	3	US-09-246-320-1	Sequence 1, Appli
	96	15.4	61.6	10398	3	US-09-243-743-1	Sequence 1, Appli
	97	15.4	61.6	10398	4	US-09-546-738-1	Sequence 1, Appli
	98	15.4	61.6	10398	4	US-09-923-726-1	Sequence 1, Appli
	99	15.4	61.6	14672	4	US-08-961-527-111	Sequence 111, App
	100	15.4	61.6	17000	4	US-09-548-797B-7	Sequence 7, Appli

C 101	15.4	61.6	32798	4	US-09-604-694B-1	Sequence 1, Appli	174	15	60.0	5086	5	PCT-US93-05851-4	Sequence 4, Appli
C 102	15.4	61.6	34303	2	US-08-735-609-4	Sequence 4, Appli	175	15	60.0	5086	5	PCT-US93-06251-2	Sequence 2, Appli
C 103	15.4	61.6	34303	2	US-08-735-609-4	Sequence 4, Appli	176	15	60.0	5094	4	US-09-234-186-7	Sequence 7, Appli
C 104	15.4	61.6	34303	3	US-09-315-372-4	Sequence 4, Appli	177	15	60.0	5104	6	5506344-1	Patent No. 5506344
C 105	15.4	61.6	34303	3	US-09-244-752-4	Sequence 4, Appli	178	15	60.0	6030	4	US-09-023-655-1015	Sequence 1015, Ap
C 106	15.4	61.6	34303	3	US-09-245-497-4	Sequence 4, Appli	179	15	60.0	44848	4	US-09-435-739-42	Sequence 42, Appli
C 107	15.4	61.6	34303	4	US-09-562-919-4	Sequence 4, Appli	180	15	60.0	536165	4	US-09-214-808-1	Sequence 1, Appli
C 108	15.4	61.6	34382	2	US-08-374-483-6	Sequence 6, Appli	181	15	60.0	786431	4	US-09-751-389-3	Sequence 3, Appli
C 109	15.4	61.6	34308	3	US-08-973-334-3	Sequence 3, Appli	182	14.8	59.2	283	4	US-09-313-294A-4190	Sequence 4190, Ap
C 110	15.4	61.6	34308	3	US-09-563-859A-3	Sequence 3, Appli	183	14.8	59.2	354	2	US-08-908-713-3	Sequence 3, Appli
C 111	15.4	61.6	34308	3	US-08-549-489-3	Sequence 3, Appli	184	14.8	59.2	1150	3	US-09-372-934-3	Sequence 3, Appli
C 112	15.4	61.6	35871	4	US-09-956-335-2	Sequence 2, Appli	185	14.8	59.2	1150	4	US-09-766-916-3	Sequence 3, Appli
C 113	15.4	61.6	35935	2	US-08-735-609-1	Sequence 1, Appli	186	14.8	59.2	1150	4	US-09-766-898-3	Sequence 3, Appli
C 114	15.4	61.6	35935	2	US-08-735-609-1	Sequence 1, Appli	187	14.8	59.2	1150	4	US-09-637-040C-3	Sequence 3, Appli
C 115	15.4	61.6	35935	3	US-08-379-452-43	Sequence 43, Appli	188	14.8	59.2	1185	4	US-09-134-001C-2117	Sequence 2117, Ap
C 116	15.4	61.6	35935	3	US-09-315-372-1	Sequence 1, Appli	189	14.8	59.2	1193	4	US-09-227-357-63	Sequence 63, Appli
C 117	15.4	61.6	35935	3	US-09-244-752-1	Sequence 1, Appli	190	14.8	59.2	1301	3	US-08-804-166-7	Sequence 7, Appli
C 118	15.4	61.6	35935	3	US-09-245-497-1	Sequence 1, Appli	191	14.8	59.2	1301	3	US-08-910-991-7	Sequence 7, Appli
C 119	15.4	61.6	35935	3	US-09-409-670-43	Sequence 43, Appli	192	14.8	59.2	1301	4	US-09-756-186-7	Sequence 7, Appli
C 120	15.4	61.6	35935	4	US-09-562-919-1	Sequence 1, Appli	193	14.8	59.2	1373	4	US-09-482-273-70	Sequence 70, Appli
C 121	15.4	61.6	35978	4	US-09-956-335-1	Sequence 1, Appli	194	14.8	59.2	1526	3	US-09-324-867-63	Sequence 63, Appli
C 122	15.4	61.6	786431	1	US-09-751-389-3	Sequence 3, Appli	195	14.8	59.2	1717	1	US-08-229-515A-12	Sequence 12, Appli
C 123	15.2	60.8	48	1	US-08-225-224-26	Sequence 26, Appli	196	14.8	59.2	1717	1	US-08-645-865-12	Sequence 12, Appli
C 124	15.2	60.8	48	3	US-08-722-258-26	Sequence 26, Appli	197	14.8	59.2	2230	4	US-09-189-527-12	Sequence 12, Appli
C 125	15.2	60.8	48	5	PCT-US93-04568-26	Sequence 26, Appli	198	14.8	59.2	2508	4	US-09-758-282B-48	Sequence 48, Appli
C 126	15.2	60.8	134	3	US-09-284-782-27	Sequence 27, Appli	199	14.8	59.2	2508	4	US-09-758-282B-242	Sequence 242, App
C 127	15.2	60.8	294	4	US-09-489-039A-3664	Sequence 3664, Ap	200	14.8	59.2	2526	4	US-09-758-282B-258	Sequence 258, App
C 128	15.2	60.8	1080	2	US-08-918-727-2	Sequence 2, Appli	201	14.8	59.2	2526	4	US-09-758-282B-262	Sequence 262, App
C 129	15.2	60.8	1080	3	US-09-205-680A-2	Sequence 2, Appli	202	14.8	59.2	2831	2	US-08-906-713-1	Sequence 1, Appli
C 130	15.2	60.8	1275	4	US-09-614-155B-4	Sequence 4, Appli	203	14.8	59.2	3158	2	US-08-464-517-36	Sequence 36, Appli
C 131	15.2	60.8	1275	4	US-09-166-963-4	Sequence 4, Appli	204	14.8	59.2	3158	2	US-08-245-361A-36	Sequence 36, Appli
C 132	15.2	60.8	1630	3	US-08-383-586-4	Sequence 4, Appli	205	14.8	59.2	3158	3	US-08-463-712-36	Sequence 36, Appli
C 133	15.2	60.8	1698	4	US-09-489-039A-3860	Sequence 3860, Ap	206	14.8	59.2	3394	4	US-09-620-312D-314	Sequence 314, App
C 134	15.2	60.8	2160	4	US-09-372-668-1	Sequence 1, Appli	207	14.8	59.2	5207	4	US-09-858-664A-1	Sequence 1, Appli
C 135	15.2	60.8	2160	4	US-09-697-377-1	Sequence 1, Appli	208	14.8	59.2	5207	4	US-10-274-978-1	Sequence 1, Appli
C 136	15.2	60.8	2160	4	US-09-696-868-1	Sequence 1, Appli	209	14.8	59.2	5683	4	US-09-845-583A-7	Sequence 7, Appli
C 137	15.2	60.8	2160	4	US-09-697-341-1	Sequence 1, Appli	210	14.8	59.2	5683	4	US-09-568-921-118	Sequence 118, App
C 138	15.2	60.8	2448	3	US-08-691-563C-53	Sequence 53, Appli	211	14.8	59.2	5702	4	US-09-576-594-14	Sequence 14, Appli
C 139	15.2	60.8	2448	4	US-08-374-766-53	Sequence 53, Appli	212	14.8	59.2	7626	4	US-08-306-691B-19	Sequence 19, Appli
C 140	15.2	60.8	2448	4	US-08-979-847B-49	Sequence 49, Appli	213	14.8	59.2	35100	1	PCT-US93-06251-19	Sequence 19, Appli
C 141	15.2	60.8	2556	4	US-08-976-259-61	Sequence 61, Appli	214	14.8	59.2	75395	5	US-09-984-890-3	Sequence 3, Appli
C 142	15.2	60.8	2675	4	US-09-636-791A-12	Sequence 12, Appli	215	14.8	59.2	75395	23	US-09-710-200-29	Sequence 29, Appli
C 143	15.2	60.8	4550	4	US-09-462-136-1	Sequence 1, Appli	216	14.6	58.4	23	4	US-09-975-408-29	Sequence 29, Appli
C 144	15.2	60.8	5474	3	US-09-040-738-1	Sequence 1, Appli	217	14.6	58.4	37	3	US-09-153-863C-33	Sequence 33, Appli
C 145	15.2	60.8	5474	3	US-08-652-426A-1	Sequence 1, Appli	218	14.6	58.4	80	2	US-07-733-095B-10	Sequence 10, Appli
C 146	15.2	60.8	2451	4	US-08-433-702B-173	Sequence 173, App	219	14.6	58.4	120	6	5510472-5	Patent No. 5510472
C 147	15.2	60.8	59065	4	US-08-813-817-3	Sequence 3, Appli	220	14.6	58.4	197	4	US-09-189-618A-15	Sequence 15, Appli
C 148	15.2	60.8	59065	4	US-09-978-197-3	Sequence 3, Appli	221	14.6	58.4	202	4	US-09-313-294A-5307	Sequence 5307, Ap
C 149	15.2	60.8	193303	4	US-09-497-855A-37	Sequence 37, Appli	222	14.6	58.4	297	4	US-09-313-294A-6906	Sequence 6906, Ap
C 150	15.2	60.8	193303	4	US-09-497-855A-44	Sequence 44, Appli	223	14.6	58.4	304	4	US-09-621-976-10937	Sequence 10937, A
C 151	15	60.0	47	4	US-09-422-978-2989	Sequence 2989, Ap	224	14.6	58.4	344	4	US-09-621-976-10937	Sequence 18897, A
C 152	15	60.0	297	4	US-08-313-294A-6878	Sequence 6878, Ap	225	14.6	58.4	351	4	US-09-621-976-10937	Sequence 361, App
C 153	15	60.0	409	4	US-08-733-234A-31	Sequence 31, Appli	226	14.6	58.4	351	4	US-09-621-976-10937	Sequence 361, App
C 154	15	60.0	431	4	US-08-312-283C-356	Sequence 356, App	227	14.6	58.4	351	4	US-09-621-976-10937	Sequence 361, App
C 155	15	60.0	507	4	US-09-105-470B-28	Sequence 28, Appli	228	14.6	58.4	351	4	US-09-621-976-10937	Sequence 361, App
C 156	15	60.0	525	4	US-09-621-976-18706	Sequence 18706, A	229	14.6	58.4	365	4	US-09-685-166A-361	Sequence 361, App
C 157	15	60.0	753	4	US-09-489-039A-210	Sequence 210, App	230	14.6	58.4	378	4	US-09-423-233-5	Sequence 5, Appli
C 158	15	60.0	1050	4	US-09-489-039A-968	Sequence 968, App	231	14.6	58.4	407	3	US-09-621-976-18899	Sequence 18899, A
C 159	15	60.0	1075	2	US-08-809-960-1	Sequence 1, Appli	232	14.6	58.4	407	3	US-09-621-976-18899	Sequence 18899, A
C 160	15	60.0	1410	2	US-08-975-316-86	Sequence 86, Appli	233	14.6	58.4	412	4	US-09-621-976-18899	Sequence 39, Appli
C 161	15	60.0	1410	4	US-08-975-316-86	Sequence 86, Appli	234	14.6	58.4	415	4	US-09-635-747-39	Sequence 40, Appli
C 162	15	60.0	1410	4	US-09-169-789-86	Sequence 86, Appli	235	14.6	58.4	415	4	US-09-635-747-39	Sequence 40, Appli
C 163	15	60.0	2172	4	US-09-484-970B-126	Sequence 126, App	236	14.6	58.4	437	4	US-09-397-787-332	Sequence 332, App
C 164	15	60.0	2461	1	US-08-282-141-1	Sequence 1, Appli	237	14.6	58.4	441	4	US-09-389-681-186	Sequence 186, App
C 165	15	60.0	2555	4	US-09-866-028-68	Sequence 68, Appli	238	14.6	58.4	441	4	US-09-620-405B-186	Sequence 186, App
C 166	15	60.0	2852	3	US-09-063-950-1	Sequence 1, Appli	239	14.6	58.4	441	4	US-09-335-338-186	Sequence 186, App
C 167	15	60.0	4825	6	5459251-1	Patent No. 5459251	240	14.6	58.4	441	4	US-09-335-338-186	Sequence 186, App
C 168	15	60.0	5086	2	US-08-465-485A-19	Sequence 19, Appli	241	14.6	58.4	441	4	US-09-433-826B-186	Sequence 186, App
C 169	15	60.0	5086	2	US-08-365-486A-14	Sequence 14, Appli	242	14.6	58.4	441	4	US-09-604-287A-186	Sequence 186, App
C 170	15	60.0	5086	3	US-09-080-285-19	Sequence 19, Appli	243	14.6	58.4	441	4	US-09-489-039A-6675	Sequence 6675, Ap
C 171	15	60.0	5086	3	US-08-880-342-14	Sequence 14, Appli	244	14.6	58.4	441	4	US-09-834-759-186	Sequence 186, App
C 172	15	60.0	5086	4	US-09-724-426-19	Sequence 19, Appli	245	14.6	58.4	459	4	US-09-489-039A-4811	Sequence 4811, Ap
C 173	15	60.0	5086	4	US-09-233-527-7	Sequence 7, Appli	246	14.6	58.4	461	4	US-09-620-312D-244	Sequence 244, App





```
US-09-146-053-7
Query Match      68.8%; Score 17.2; DB 4; Length 16595;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGGCACATAGC 25
Db 12850 GGAGGGCCACAGGGCCCAAGC 12871

RESULT 4
US-09-621-976-167/c
Sequence 167, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 167
LENGTH: 481
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 297..479
NAME/KEY: sig_peptide
LOCATION: 297..353
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 5.4000000536743
OTHER INFORMATION: seq LCCSGCVPSLCCS/SY
US-09-621-976-167

Query Match      67.2%; Score 16.8; DB 4; Length 481;
Best Local Similarity 81.8%; Pred. No. 1.7e+02;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGGCACATAGC 25
Db 376 GGAGCAACAGAGGGSACATAGC 355

RESULT 5
US-09-604-978-5/c
Sequence 5, Application US/09604978
Patent No. 6455674
GENERAL INFORMATION:
APPLICANT: Einat, Paz
APPLICANT: Skaliter, Rami
TITLE OF INVENTION: HYPOXIA-REGULATED GENES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: KOHN & ASSOCIATES
STREET: 30500 No. 6455674thwestern Hwy., Suite 401
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/604,978
FILING DATE: 28-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/138,112
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 0168.00034
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 5395055
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2059 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-09-146-053-7
Query Match      67.2%; Score 16.8; DB 4; Length 2059;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 AGGGCCAGAGGGGCACATAGC 25
Db 154 AGGGCCAGAGGGGCACATAGC 135

RESULT 6
US-09-604-728-5/c
Sequence 5, Application US/09604728
Patent No. 6555667
GENERAL INFORMATION:
APPLICANT: Einat, Paz
APPLICANT: Skaliter, Rami
TITLE OF INVENTION: HYPOXIA-REGULATED GENES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: KOHN & ASSOCIATES
STREET: 30500 No. 6555667thwestern Hwy., Suite 401
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/604,728
FILING DATE: 28-Jun-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/138,112
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,955
REFERENCE/DOCKET NUMBER: 0168.00034
TELECOMMUNICATION INFORMATION:
TELEPHONE: (248) 539-5050
TELEFAX: (248) 5395055
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2059 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
```

US-09-604-728-5  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 5;  
Query Match 67.2%; Score 16.8; DB 4; Length 2059;  
Best Local Similarity 90.0%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 AGGCCAGGCGGCACATAGC 25  
Db 154 AGGCCAGGCGGCACAGC 135

RESULT 7  
US-08-049-283A-1/c  
; Sequence 1, Application US/08049283A  
; Patent No. 5502176  
; GENERAL INFORMATION:  
; APPLICANT: Tenen, Daniel G.  
; APPLICANT: Pahl, Heike L.  
; APPLICANT: Burn, Timothy C.  
; TITLE OF INVENTION: Cell Specific Promoter and Uses Thereof  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/049,283A  
; FILING DATE: 14-APR-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/020,465  
; FILING DATE: 19-FEB-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/837,776  
; FILING DATE: 13-FEB-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brook, David E.  
; REGISTRATION NUMBER: 22,592  
; REFERENCE/DOCKET NUMBER: BIH91-03'A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 533 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-049-283A-1

Query Match 66.4%; Score 16.6; DB 1; Length 533;  
Best Local Similarity 82.6%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CGGAGGCGGCAGGCGGCATAGC 25  
Db 351 CGGAGGCTCAGTGAGGCATAGC 329

RESULT 8  
US-08-618-100B-3  
; Sequence 3, Application US/08618100B  
; Patent No. 6068976  
; GENERAL INFORMATION:  
; APPLICANT: Briggs, Michael R.  
; APPLICANT: Auwerx, Johan  
; APPLICANT: de Vos, Piet  
; APPLICANT: Staels, Bart  
; APPLICANT: Croston, Glenn E.  
; APPLICANT: Miller, Stephen G.  
; TITLE OF INVENTION: MODULATORS OF ob GENE AND  
; TITLE OF INVENTION: SCREENING METHODS THEREFOR  
; NUMBER OF SEQUENCES: 48  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FastSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/618,100B  
; FILING DATE: March 19, 1996  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/558,588  
; FILING DATE: October 30, 1995  
; APPLICATION NUMBER: 08/510,584  
; FILING DATE: August 2, 1995  
; APPLICATION NUMBER: 08/418,096  
; FILING DATE: April 5, 1995  
; APPLICATION NUMBER: 08/408,584  
; FILING DATE: March 20, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 219/075  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10684 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; DESCRIPTION: Sequence between exon 1 and exon 2  
; Patent No. 6068976  
US-08-618-100B-3

Query Match 66.4%; Score 16.6; DB 3; Length 10684;  
Best Local Similarity 82.6%; Pred. No. 2.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CCGAGGCGGCAGGCGGCATAG 24  
Db 2252 CCTGGTGCTAGAGGCGCATAG 2274

RESULT 9  
US-09-277-262-3/c  
; Sequence 3, Application US/09277262  
; Patent No. 6395482

```
/ GENERAL INFORMATION:
/ APPLICANT: Karayiorougou, Maria
/ APPLICANT: Gogos, Joseph A
/ TITLE OF INVENTION: METHODS OF DETERMINING A SUSCEPTIBILITY TO OR PRESENCE
/ TITLE OF INVENTION: OF SCHIZOPHRENIA, OR A DISEASE OR DISORDER RELATED
/ TITLE OF INVENTION: THEREO
/ FILE REFERENCE: 600-1-223 CIP
/ CURRENT APPLICATION NUMBER: US/09/277,262
/ CURRENT FILING DATE: 1999-03-26
/ EARLIER APPLICATION NUMBER: 09/229,530
/ EARLIER FILING DATE: 1999-01-13
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 3
/ LENGTH: 1494
/ TYPE: DNA
/ ORGANISM: Mus musculus
US-09-277-262-3

Query Match      65.6%; Score 16.4; DB 4; Length 1494;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      6 AGGGCCAGAGGGCACATA 23
DB      1350 AGGGCCATAGGGCACATA 1333

RESULT 10
US-09-277-262-7/c
/ Sequence 7, Application US/09277262
/ Patent No. 6395482
/ GENERAL INFORMATION:
/ APPLICANT: Karayiorougou, Maria
/ APPLICANT: Gogos, Joseph A
/ TITLE OF INVENTION: METHODS OF DETERMINING A SUSCEPTIBILITY TO OR PRESENCE
/ TITLE OF INVENTION: OF SCHIZOPHRENIA, OR A DISEASE OR DISORDER RELATED
/ TITLE OF INVENTION: THEREO
/ FILE REFERENCE: 600-1-223 CIP
/ CURRENT APPLICATION NUMBER: US/09/277,262
/ CURRENT FILING DATE: 1999-03-26
/ EARLIER APPLICATION NUMBER: 09/229,530
/ EARLIER FILING DATE: 1999-01-13
/ NUMBER OF SEQ ID NOS: 9
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 7
/ LENGTH: 2240
/ TYPE: DNA
/ ORGANISM: Mus musculus
US-09-277-262-7

Query Match      65.6%; Score 16.4; DB 4; Length 2240;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      6 AGGGCCAGAGGGCACATA 23
DB      1646 AGGGCCATAGGGCACATA 1629

RESULT 11
US-09-482-273-44
/ Sequence 44, Application US/09482273
/ Patent No. 6534631
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: 71 Human Secreted Proteins
/ FILE REFERENCE: P2030P1
/ CURRENT APPLICATION NUMBER: US/09/482,273
/ CURRENT FILING DATE: 2000-01-13
/ EARLIER APPLICATION NUMBER: PCT/US99/15849
/ EARLIER FILING DATE: 1999-07-14
/ EARLIER APPLICATION NUMBER: 60/092,921
```

```
/ EARLIER FILING DATE: 1998-07-15
/ EARLIER APPLICATION NUMBER: 60/092,922
/ EARLIER FILING DATE: 1998-07-15
/ EARLIER APPLICATION NUMBER: 60/092,956
/ EARLIER FILING DATE: 1998-07-15
/ NUMBER OF SEQ ID NOS: 267
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 44
/ LENGTH: 1262
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-482-273-44

Query Match      64.8%; Score 16.2; DB 4; Length 1262;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 CCGAGGGCCAGAGGGCACAT 22
DB      1004 CCTAAGGGCCAGAGGGCACAT 1024

RESULT 12
US-09-448-176-9
/ Sequence 9, Application US/09448176
/ Patent No. 6248533
/ GENERAL INFORMATION:
/ APPLICANT: KAMIZONO, Shintaro
/ APPLICANT: YAMADA, Akira
/ APPLICANT: HIGUCHI, Takafumi
/ APPLICANT: KATO, Hirohisa
/ APPLICANT: ITOH, Kyogo
/ APPLICANT: SEKI, Naoko
/ TITLE OF INVENTION: GENE DIAGNOSIS OF DISEASES WHEREIN
/ TITLE OF INVENTION: TNF-ALPHA PROMOTERS PARTICIPATE
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: FOLEY & LARDNER
/ STREET: 3000 K Street, N.W.
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20007-5109
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/448,176
/ FILING DATE: 24-No. 6248533-1999
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: WO PCT/JP98/02270
/ FILING DATE: 25-MAY-1998
/ APPLICATION NUMBER: WO PCT/JP97/04304
/ FILING DATE: 26-NOV-1997
/ APPLICATION NUMBER: JP 9-173900
/ FILING DATE: 30-JUN-1997
/ APPLICATION NUMBER: JP 9-134973
/ FILING DATE: 26-MAY-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Wegner, Harold C.
/ REGISTRATION NUMBER: 25,258
/ REFERENCE/DOCKET NUMBER: 74129/472
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 672-5300
/ TELEFAX: (202) 672-5399
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1357 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
```

TOPLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 9;  
US-09-448-176-9

Query Match 64.8%; Score 16.2; DB 3; Length 1357;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGCACATAG 24  
|||||  
DB 56 GGAGGGACAGAGGGCTCAAAG 76

RESULT 14  
US-09-448-176-11  
; Sequence 11, Application US/09448176  
; Patent No. 6248533  
; GENERAL INFORMATION:  
; APPLICANT: KAMIZONO, Shintaro  
; YAMADA, Akira  
; HIGUCHI, Takafumi  
; KATO, Hirochisa  
; ITOH, Kyogo  
; SEKI, Naoko  
; TITLE OF INVENTION: GENE DIAGNOSIS OF DISEASES WHEREIN  
; TNF-ALPHA PROMOTERS PARTICIPATE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/448,176  
; FILING DATE: 24-NO. 6248533-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP98/02270  
; FILING DATE: 25-MAY-1998  
; APPLICATION NUMBER: WO PCT/JP97/04304  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: JP 9-173900  
; FILING DATE: 30-JUN-1997  
; APPLICATION NUMBER: JP 9-134973  
; FILING DATE: 26-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wegner, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 74129/472  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1357 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
US-09-448-176-11

Query Match 64.8%; Score 16.2; DB 3; Length 1357;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGCACATAG 24  
|||||  
DB 56 GGAGGGACAGAGGGCTCAAAG 76

RESULT 15  
US-08-762-428A-5/c

TOPLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 9;  
US-09-448-176-9

Query Match 64.8%; Score 16.2; DB 3; Length 1357;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GGAGGGCCAGAGGGCACATAG 24  
|||||  
DB 56 GGAGGGACAGAGGGCTCAAAG 76

RESULT 13  
US-09-448-176-10  
; Sequence 10, Application US/09448176  
; Patent No. 6248533  
; GENERAL INFORMATION:  
; APPLICANT: KAMIZONO, Shintaro  
; YAMADA, Akira  
; HIGUCHI, Takafumi  
; KATO, Hirochisa  
; ITOH, Kyogo  
; SEKI, Naoko  
; TITLE OF INVENTION: GENE DIAGNOSIS OF DISEASES WHEREIN  
; TNF-ALPHA PROMOTERS PARTICIPATE  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/448,176  
; FILING DATE: 24-NO. 6248533-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/JP98/02270  
; FILING DATE: 25-MAY-1998  
; APPLICATION NUMBER: WO PCT/JP97/04304  
; FILING DATE: 26-NOV-1997  
; APPLICATION NUMBER: JP 9-173900  
; FILING DATE: 30-JUN-1997  
; APPLICATION NUMBER: JP 9-134973  
; FILING DATE: 26-MAY-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wegner, Harold C.  
; REGISTRATION NUMBER: 25,258  
; REFERENCE/DOCKET NUMBER: 74129/472  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1357 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-448-176-10

Query Match 64.8%; Score 16.2; DB 3; Length 1357;  
Best Local Similarity 85.7%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

/ Sequence 5, Application US/08762428A  
/ Patent No. 6120993  
/ GENERAL INFORMATION:  
/ APPLICANT: Ye, Guo-jie  
/ APPLICANT: Breslow, Esther M.  
/ APPLICANT: Meister, Alton  
/ TITLE OF INVENTION: 5-OXOPROLINASE  
/ NUMBER OF SEQUENCES: 10  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
/ STREET: Clinton Square, P.O. Box 1051  
/ CITY: Rochester  
/ STATE: New York  
/ COUNTRY: USA  
/ ZIP: 14603  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patent in Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/762,428A  
/ FILING DATE:  
/ CLASSIFICATION: 536  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: TIMIAN, SUSAN J.  
/ REGISTRATION NUMBER: 34,103  
/ REFERENCE/DOCKET NUMBER: 19603/1330  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 716-263-1636  
/ TELEFAX: 716-263-1600  
/ INFORMATION FOR SEQ ID NO: 5:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 3867 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: CDNA  
/ US-08-762-428A-5

Query Match 64.8%; Score 16.2; DB 3; Length 3867;  
Best Local Similarity 85.7%; Pred. No. 3.2e-02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGCACA 21  
DB 875 CCAGAGGGCCAGAGGCACA 855

RESULT 16  
US-08-762-428A-7/c  
/ Sequence 7, Application US/08762428A  
/ Patent No. 6120993  
/ GENERAL INFORMATION:  
/ APPLICANT: Ye, Guo-jie  
/ APPLICANT: Breslow, Esther M.  
/ APPLICANT: Meister, Alton  
/ TITLE OF INVENTION: 5-OXOPROLINASE  
/ NUMBER OF SEQUENCES: 10  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP  
/ STREET: Clinton Square, P.O. Box 1051  
/ CITY: Rochester  
/ STATE: New York  
/ COUNTRY: USA  
/ ZIP: 14603  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patent in Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/762,428A

/ FILING DATE:  
/ CLASSIFICATION: 536  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: TIMIAN, SUSAN J.  
/ REGISTRATION NUMBER: 34,103  
/ REFERENCE/DOCKET NUMBER: 19603/1330  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 716-263-1636  
/ TELEFAX: 716-263-1600  
/ INFORMATION FOR SEQ ID NO: 7:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 4016 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: CDNA  
/ US-08-762-428A-7

Query Match 64.8%; Score 16.2; DB 3; Length 4016;  
Best Local Similarity 85.7%; Pred. No. 3.2e-02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGGAGGGCCAGAGGCACA 21  
DB 980 CCAGAGGGCCAGAGGCACA 960

RESULT 17  
US-09-050-159-123/c  
/ Sequence 123, Application US/09050159A  
/ Patent No. 6197505  
/ GENERAL INFORMATION:  
/ APPLICANT: No. 6197505berg, Leif T  
/ APPLICANT: Andersson, Maria K  
/ APPLICANT: Linstrom, Per H  
/ TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND  
/ TITLE OF INVENTION: COMPOSITIONS FOR USE THEREOF  
/ FILE REFERENCE: 1248/1D042  
/ CURRENT APPLICATION NUMBER: US/09/050,159A  
/ EARLIER FILING DATE: 1998-03-27  
/ EARLIER FILING DATE: 1987-04-03  
/ NUMBER OF SEQ ID NOS: 133  
/ SOFTWARE: Patent in Ver. 2.1  
/ SEQ ID NO 123  
/ LENGTH: 1278  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ OTHER INFORMATION: Angiotensinogen, 5' region and exon 1  
/ US-09-050-159-123

Query Match 64.0%; Score 16; DB 3; Length 1278;  
Best Local Similarity 79.2%; Pred. No. 3.8e-02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CCGGAGGGCCAGAGGCACATAGC 25  
DB 840 CCGGAGGTGCAGAGGGCCAGAGGGC 817

RESULT 18  
US-09-976-594-193  
/ Sequence 193, Application US/09976594  
/ Patent No. 6673549  
/ GENERAL INFORMATION:  
/ APPLICANT: Furness, Michael  
/ APPLICANT: Buchbinder, Jenny  
/ TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
/ FILE REFERENCE: PA-0041 US  
/ CURRENT APPLICATION NUMBER: US/09/976,594  
/ CURRENT FILING DATE: 2001-10-12

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 193
; LENGTH: 1973
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 235356.1
; NAME/KEY: unsure
; LOCATION: 614
; OTHER INFORMATION: a, t, c, g, or other
US-09-376-594-193

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 1973;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CCGGAGGGCCAGAGGGCCACATAG 25
Db 1437 CCTGAGCCCGAGAGTGGACACAGC 1460

RESULT 19
US-09-566-921-52/c
; Sequence 52 Application US/09566921
; Patent No. 6682888
; GENERAL INFORMATION:
; APPLICANT: Loring, Jeanne F.
; APPLICANT: Tingley, Debora W.
; APPLICANT: Edwards, Carla M.
; TITLE OF INVENTION: GENES EXPRESSED IN ALZHEIMER'S DISEASE
; FILE REFERENCE: PA-0024 US
; CURRENT APPLICATION NUMBER: US/09/566,921
; CURRENT FILING DATE: 2000-05-05
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PERL Program
; SEQ ID NO 52
; LENGTH: 2201
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6682888 223035.8
US-09-566-921-52

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 2201;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCCACATAG 24
Db 111 CCCCTGGGGCCAGAGGGCCACATGG 88

RESULT 20
US-09-462-917A-106
; Sequence 106 Application US/09462917A
; Patent No. 6511801
; GENERAL INFORMATION:
; APPLICANT: Delaporte, Eric
; APPLICANT: Peeters, Martine
; APPLICANT: Saman, Eric
; APPLICANT: Vanden Haesevelde, Marlen
; TITLE OF INVENTION: HIV-1 Group O Antigens and Uses Thereof
; FILE REFERENCE: INNS:014 11362.0014.NPUS00
; CURRENT APPLICATION NUMBER: US/09/462,917A
; CURRENT FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: PCT/EP98/04522
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: Patentin version 3.0

; SEQ ID NO 106
; LENGTH: 3600
; TYPE: DNA
; ORGANISM: Human
US-09-462-917A-106

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 3600;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCCACATAG 24
Db 797 CCCAGAGACCAGGGGGCCACAGAG 820

RESULT 21
US-08-943-731-4/c
; Sequence 4 Application US/08943731
; Patent No. 6265157
; GENERAL INFORMATION:
; APPLICANT: PROCKOP, DARWIN J.
; APPLICANT: SPOTILA, LORETTA D.
; APPLICANT: DELTAS, CONSTANTINOS D.
; APPLICANT: SEREDA, LARISA W.
; APPLICANT: LARSON, ANDREA W.
; APPLICANT: PACK, MICHAEL
; APPLICANT: COLIGE, ALAIN
; APPLICANT: EARLY, JAMES
; APPLICANT: KORKKO, JARMO
; APPLICANT: ALA-KOKKO, LEENA, et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STREET: FLR
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,731
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,322
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,628
; FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: DOYLE LEARY Ph.D., KATHRYN
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9599-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17606 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-943-731-4

Query Match 64.0%; Score 16; DB 3; Length 17606;  
Best Local Similarity 79.2%; Pred. No. 4e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels

**QY**            1 CCGGAGGGCCACAGGGCACATAG 24  
             ||| ||||| ||||| |||||  
**Dδ**            15602 CCCGAGGGCCACAGGTGCCATGG 15579

RESULT 22

```

US-09-015-188-5/C
; Sequence 5, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 5
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: "n" at all locations is "unknown"
US-09-015-188-5

```

Query Match 63.2%; Score 15.8; DB 4; Length 1231;  
Best Local Similarity 77.3%; Pred. No. 4.6e+02;  
Matches 17; Conservative 0; Mismatches 5; Indels 0

QY  
db

1 CCCGAGGGCCAGAGGCACAT 22  
643 CCGGGNGGCCAGAGGGNACNT 622

## RESULT 23

```

US-09-210-748A-3
; Sequence 3, Application US/09210748A
; Patent No. 6335156
; GENERAL INFORMATION:
; APPLICANT: Hermeking, Heiko
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: 14-3-3 SIGMA ARREST THE CELL CYCLE
; FILE REFERENCE: 1107.77810
; CURRENT APPLICATION NUMBER: US/09/210,748A
; CURRENT FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/069,415
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 7680
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-210-748A-3

```

Query Match 63.2%; Score 15.8; DB 4; Length 7680;  
Best Local Similarity 89.5%; Pred. No. 4.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0

Qy 4 GGAGGGCCAGAGGGCACAT 22  
7601 GGAGGGCCAGAGGGTAAAT 7619  
db

RESULT 24

US-08-764-233A-1/c

```

/ Sequence 1, Application US/08764233A
/ Patent No. 5716849
/ GENERAL INFORMATION:
/ APPLICANT: Ligon, James M.
/ APPLICANT: Schupp, Thomas
/ APPLICANT: Beck, James J.
/ APPLICANT: Hill, Dwight S.
/ APPLICANT: Neff, Snezana
/ APPLICANT: Ryals, John A.
/ TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Ciba-Geigy Corporation
/ STREET: 520 White Plains Road, P.O. Box 2005
/ CITY: Tarrytown
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10591
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/764,233A
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/729,214
/ FILING DATE: 09-OCT-1996
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/258,261
/ FILING DATE: 08-JUN-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: McIgis, J. Timothy
/ REGISTRATION NUMBER: 38,241
/ REFERENCE/DOCKET NUMBER: 1506/CIP6
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (919) 541-8587
/ TELEFAX: (919) 541-8689
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 49377 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: Sorangium cellulosum
/ CLONE: p98/1, pJL3, and pVKM15
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 383..760
/ OTHER INFORMATION: /product= "SorR"
/ OTHER INFORMATION: /note= "This gene encodes a protein th
/ OTHER INFORMATION: the reductase domains of type I PKSs s
/ OTHER INFORMATION: Saccharopolyspora erythraea."
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 927..19874
/ OTHER INFORMATION: /product= "SorA"
/ OTHER INFORMATION: /note= "Gene product is highly homolog
/ OTHER INFORMATION: are known to be involved in the synthe
/ OTHER INFORMATION: compounds."
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 942..7115
/ OTHER INFORMATION: /product= "Module 1 of SorA".
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 7203..12884
/ OTHER INFORMATION: /product= "Module 2 of SorA"

```



FEATURE:  
NAME/KEY: misc feature  
LOCATION: 13455..19616  
OTHER INFORMATION: /product= "Module 3 of SorA"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19871..46318  
OTHER INFORMATION: /product= "SorB"  
OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS genes"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 19870..24556  
OTHER INFORMATION: /product= "Module 1 of SorB"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 24638..30820  
OTHER INFORMATION: /product= "Module 2 of SorB"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 30881..35446  
OTHER INFORMATION: /product= "Module 3 of SorB"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 35528..40114  
OTHER INFORMATION: /product= "Module 4 of SorB"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 40190..46318  
OTHER INFORMATION: /product= "Module 5 of SorB"  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 46851..47891  
OTHER INFORMATION: /product= "SorM"  
OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly homologous to the methyltransferase from Streptomyces hygroscopicus that is involved in the synthesis of the polyketide rapamycin."  
US-08-764-233A-1  
Query Match 63.2%; Score 15.8; DB 1; Length 49377;  
Best Local Similarity 89.5%; Pred. No. 5e+02; Mismatches 0; Gaps 0;  
Matches 17; Conservative 0;  
QY 1 CCCGAGGGCCAGAGGGCA 19  
DB 15287 CCTGAGGCCAGAGGGCA 15269  
RESULT 25  
US-09-621-976-1421/c  
Sequence 1421, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET.054PR2  
CURRENT APPLICATION NUMBER: US/09/621.976  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 1421  
LENGTH: 539  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 243..467  
NAME/KEY: sig\_peptide  
LOCATION: 243..437  
OTHER INFORMATION: Von Heijne matrix  
OTHER INFORMATION: score 8.19999980926514

OTHER INFORMATION: seq LEFLIOLVTVARG/CW  
US-09-621-976-1421  
Query Match 62.4%; Score 15.6; DB 4; Length 539;  
Best Local Similarity 81.8%; Pred. No. 5.4e+02; Mismatches 18; Conservative 0; Indels 0; Gaps 0;  
QY 4 GGAGGCCAGAGGGCACATAGC 25  
DB 381 GGAGCAACAGAGGGACATAGC 360  
RESULT 26  
US-09-719-919A-2/c  
Sequence 2, Application US/09719919A  
Patent No. 6680180  
GENERAL INFORMATION:  
APPLICANT: Jegla, Timothy James  
APPLICANT: ICAGEN Incorporated  
TITLE OF INVENTION: Kv6.2, a Voltage-Gated Potassium Channel Subunit  
FILE REFERENCE: 018512-001410US  
CURRENT APPLICATION NUMBER: US/09/719.919A  
CURRENT FILING DATE: 2001-02-22  
PRIOR APPLICATION NUMBER: US 60/091,466  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: WO PCT/US99/14945  
PRIOR FILING DATE: 1999-06-30  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 1518  
TYPE: DNA  
ORGANISM: Mus sp.  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1518)  
OTHER INFORMATION: mouse alpha subunit of heteromeric voltage-gated potassium channel Kv6.2  
US-09-719-919A-2  
Query Match 62.4%; Score 15.6; DB 4; Length 1518;  
Best Local Similarity 81.8%; Pred. No. 5.6e+02; Mismatches 18; Conservative 0; Indels 0; Gaps 0;  
QY 1 CCCGAGGGCCAGAGGGCACAT 22  
DB 430 CCCGAGGACAGAGGGCACAT 409  
RESULT 27  
US-09-061-154-1/c  
Sequence 1, Application US/09061154  
Patent No. 6462188  
GENERAL INFORMATION:  
APPLICANT: Kirkness, E.  
TITLE OF INVENTION: NEW HUMAN 5-HT3 RECEPTOR  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 2000 Pennsylvania Avenue, NW, suite 5500  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20006-1888  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/061,154  
FILING DATE: 16-APR-1998  
CLASSIFICATION: 536

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24366-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 67...1389
; OTHER INFORMATION:
; US-09-061-154-1
;
; Query Match 62.4%; Score 15.6; DB 4; Length 1613;
; Best Local Similarity 81.8%; Pred. No. 5.6e-02;
; Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 3 CGGAGGGCCGAGGCGACATAG 24
; DB 415 CGGGGGCCGAGATGGCACTTAG 394
;
; RESULT 28
; US-09-061-154-3/c
; Sequence 3, Application US/09061154
; Patent No. 6462188
; GENERAL INFORMATION:
; APPLICANT: Kirkness, E.
; TITLE OF INVENTION: NEW HUMAN 5-HT3 RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/061,154
; FILING DATE: 16-APR-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24366-20001.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 67...1389
; OTHER INFORMATION:
; US-09-061-154-1
;
; Query Match 62.4%; Score 15.6; DB 4; Length 1613;
; Best Local Similarity 81.8%; Pred. No. 5.6e-02;
; Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 3 CGGAGGGCCGAGGCGACATAG 24
; DB 415 CGGGGGCCGAGATGGCACTTAG 394
;
; RESULT 29
; US-08-991-426-3
; Sequence 3, Application US/08991426
; Patent No. 6013257
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NEUROTACTIN AND USES THEREFOR
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/991,426
; FILING DATE: 16-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/851,160
; FILING DATE: 05-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/643,798
; FILING DATE: 07-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/043001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1654 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 86...1276
; OTHER INFORMATION:
; US-08-991-426-3
;
; Query Match 62.4%; Score 15.6; DB 3; Length 1654;
; Best Local Similarity 81.8%; Pred. No. 5.6e-02;
; Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 4 CGAGGGCCGAGGCGACATAG 25
; DB 617 GGAGGGCCCTGTGGGCGAGGC 638
;
; US-10-624-714-17.rni
```

```
RESULT 30
US-09-143-470-3
; Sequence 3, Application US/09143470
; Patent No. 6043086
; GENERAL INFORMATION:
; APPLICANT: Pal, Yang
; TITLE OF INVENTION: NEUROTACTIN AND USES THEREFOR
; FILE REFERENCE: 09404/049001
; CURRENT APPLICATION NUMBER: US/09/143,470
; CURRENT FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1654
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (86)...(1276)
US-09-143-470-3
Query Match      62.4%; Score 15.6; DB 3; Length 1654;
Best Local Similarity 81.8%; Pred. No. 5.6e+02;
Matches 18, Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      4 GGAGGGCCAGAGGGCACATAGC 25
        |||||
Db      617 GGAGGGCCTGTGGCAGGAGC 638
```

Search completed: June 20, 2004, 11:47:06  
Job time : 39.7705 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 141.402 Seconds  
(without alignments)  
809.940 Million cell updates

Title: US-10-624-714-17

Perfect score:

Sequence: 1 cccgagggccagagggcacatagc 25

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs. 2290544650 residues.

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 2000000000

Post-occupancy evaluation: Minimizing the negative impact of the COVID-19 pandemic on the built environment

Post-processing:	Minimum	Match	0%
	Maximum	Match	100%

Maximum Match 100%  
Listing first 300 summaries

Database : Published Applications NA:\*

- ```

1: /cgm2_6/ptodata/2/pubnpa/US07_PUBCOMB.seq.*
2: /cgm2_6/ptodata/2/pubnpa/PCT_NEW_PUB.seq.*
3: /cgm2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq.*
4: /cgm2_6/ptodata/2/pubnpa/US06_PUBCOMB.seq.*
5: /cgm2_6/ptodata/2/pubnpa/US07_NEW_PUB.seq.*
6: /cgm2_6/ptodata/2/pubnpa/PCTUS_PUBCOMB.seq.*
7: /cgm2_6/ptodata/2/pubnpa/US08_NEW_PUB.seq.*
8: /cgm2_6/ptodata/2/pubnpa/US08_PUBCOMB.seq.*
9: /cgm2_6/ptodata/2/pubnpa/US09A_PUBCOMB.seq.*
10: /cgm2_6/ptodata/2/pubnpa/US09B_PUBCOMB.seq.*
11: /cgm2_6/ptodata/2/pubnpa/US09C_PUBCOMB.seq.*
12: /cgm2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq.*
13: /cgm2_6/ptodata/2/pubnpa/US09_NEW_PUB.seq.*
14: /cgm2_6/ptodata/2/pubnpa/US10A_PUBCOMB.seq.*
15: /cgm2_6/ptodata/2/pubnpa/US10B_PUBCOMB.seq.*
16: /cgm2_6/ptodata/2/pubnpa/US10C_PUBCOMB.seq.*
17: /cgm2_6/ptodata/2/pubnpa/US10_NEW_PUB.seq.*
18: /cgm2_6/ptodata/2/pubnpa/US06_NEW_PUB.seq.*
19: /cgm2_6/ptodata/2/pubnpa/US06_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query |        | DB | ID                   | Description        |
|------------|-------|-------|--------|----|----------------------|--------------------|
|            |       | Match | %      |    |                      |                    |
| C 1        | 25    | 100.0 | 1584   | 9  | US-09-712-363-10     | Sequence 10, Appl  |
| C 2        | 20.2  | 80.8  | 276820 | 13 | US-10-271-416-9      | Sequence 9, Appl   |
| C 3        | 18.8  | 75.2  | 813    | 13 | US-10-027-532-144075 | Sequence 144075,   |
| C 4        | 18.8  | 75.2  | 813    | 16 | US-10-027-532-144075 | Sequence 144075,   |
| C 5        | 18.6  | 74.4  | 59817  | 12 | US-10-052-482-190    | Sequence 190, App  |
| C 6        | 18.4  | 73.6  | 14546  | 10 | US-09-960-706-1043   | Sequence 1043, App |
| C 7        | 18.4  | 73.6  | 14546  | 10 | US-09-873-319-691    | Sequence 691, App  |
| C 8        | 18.2  | 72.8  | 90442  | 15 | US-10-105-637-1      | Sequence 1, Appl   |
| C 9        | 18.2  | 72.8  | 90442  | 16 | US-10-034-550-43     | Sequence 43, Appl  |
| C 10       | 17.6  | 70.4  | 523    | 15 | US-10-029-386-11131  | Sequence 11131, P  |
| C 11       | 17.6  | 70.4  | 637    | 10 | US-09-791-279-49     | Sequence 49, App   |
| C 12       | 17.6  | 70.4  | 834    | 13 | US-10-027-532-140370 | Sequence 140370,   |
| C 13       | 17.6  | 70.4  | 834    | 16 | US-10-027-532-140370 | Sequence 140370,   |
| C 14       | 17.6  | 70.4  | 810    | 13 | US-10-027-532-251702 | Sequence 251702,   |
| C 15       | 17.6  | 70.4  | 1034   | 13 | US-10-027-532-251702 | Sequence 251702,   |

|     |      |      |         |    |                      |                    |    |     |    |      |      |    |                      |                    |
|-----|------|------|---------|----|----------------------|--------------------|----|-----|----|------|------|----|----------------------|--------------------|
| 88  | 16.8 | 67.2 | 13240   | 16 | US-10-292-798-1333   | Sequence 1333, Ap  | 16 | 161 | 16 | 64.0 | 323  | 16 | US-10-242-535A-6708  | Sequence 6708, Ap  |
| 89  | 16.8 | 67.2 | 13830   | 15 | US-10-017-161-1581   | Sequence 1581, Ap  | 16 | 162 | 16 | 64.0 | 431  | 9  | US-09-983-965-4592   | Sequence 4592, Ap  |
| 90  | 16.8 | 67.2 | 13830   | 16 | US-10-292-798-1257   | Sequence 1257, Ap  | 16 | 163 | 16 | 64.0 | 438  | 9  | US-09-864-761-11-5   | Sequence 5, Appl   |
| 91  | 16.8 | 67.2 | 13830   | 16 | US-10-292-798-1255   | Sequence 1255, Ap  | 16 | 164 | 16 | 64.0 | 444  | 9  | US-09-864-761-3157   | Sequence 3157, Ap  |
| 92  | 16.8 | 67.2 | 87415   | 16 | US-10-087-192-1651   | Sequence 1651, Ap  | 16 | 165 | 16 | 64.0 | 459  | 13 | US-10-085-783A-43067 | Sequence 43067, A  |
| 93  | 16.6 | 66.4 | 364     | 9  | US-09-925-297-376    | Sequence 376, App  | 16 | 166 | 16 | 64.0 | 459  | 16 | US-10-242-535A-43067 | Sequence 43067, A  |
| 94  | 16.6 | 66.4 | 700     | 13 | US-10-027-632-162875 | Sequence 162875, A | 16 | 167 | 16 | 64.0 | 460  | 10 | US-09-918-995-28156  | Sequence 28156, A  |
| 95  | 16.6 | 66.4 | 700     | 16 | US-10-027-632-162875 | Sequence 162875, A | 16 | 168 | 16 | 64.0 | 460  | 10 | US-10-060-036-3048   | Sequence 3048, Ap  |
| 96  | 16.6 | 66.4 | 725     | 13 | US-10-027-632-103842 | Sequence 103842, A | 16 | 169 | 16 | 64.0 | 473  | 13 | US-10-085-783A-35519 | Sequence 35519, A  |
| 97  | 16.6 | 66.4 | 725     | 13 | US-10-027-632-323554 | Sequence 323554, A | 16 | 170 | 16 | 64.0 | 473  | 13 | US-10-085-783A-35519 | Sequence 35519, A  |
| 98  | 16.6 | 66.4 | 725     | 16 | US-10-027-632-103842 | Sequence 103842, A | 16 | 171 | 16 | 64.0 | 490  | 9  | US-09-925-297-1172   | Sequence 172, App  |
| 99  | 16.6 | 66.4 | 725     | 16 | US-10-027-632-323554 | Sequence 323554, A | 16 | 172 | 16 | 64.0 | 490  | 15 | US-10-060-036-3697   | Sequence 3697, Ap  |
| 100 | 16.6 | 66.4 | 761     | 13 | US-10-027-632-30195  | Sequence 30195, A  | 16 | 173 | 16 | 64.0 | 516  | 10 | US-09-770-961-116    | Sequence 116, App  |
| 101 | 16.6 | 66.4 | 761     | 16 | US-10-027-632-30195  | Sequence 30195, A  | 16 | 174 | 16 | 64.0 | 525  | 15 | US-10-029-386-20660  | Sequence 20660, A  |
| 102 | 16.6 | 66.4 | 765     | 13 | US-10-027-632-104151 | Sequence 104151, A | 16 | 175 | 16 | 64.0 | 528  | 12 | US-09-963-034-917    | Sequence 917, App  |
| 103 | 16.6 | 66.4 | 765     | 13 | US-10-027-632-104151 | Sequence 104151, A | 16 | 176 | 16 | 64.0 | 541  | 15 | US-10-029-386-3050   | Sequence 9050, Ap  |
| 104 | 16.6 | 66.4 | 765     | 13 | US-10-027-632-323670 | Sequence 323670, A | 16 | 177 | 16 | 64.0 | 555  | 13 | US-10-029-386-22402  | Sequence 22402, A  |
| 105 | 16.6 | 66.4 | 765     | 13 | US-10-027-632-323670 | Sequence 323670, A | 16 | 178 | 16 | 64.0 | 555  | 15 | US-10-029-386-24654  | Sequence 24654, A  |
| 106 | 16.6 | 66.4 | 765     | 16 | US-10-027-632-103927 | Sequence 103927, A | 16 | 179 | 16 | 64.0 | 567  | 13 | US-10-029-386-24654  | Sequence 24654, A  |
| 107 | 16.6 | 66.4 | 765     | 16 | US-10-027-632-104151 | Sequence 104151, A | 16 | 180 | 16 | 64.0 | 567  | 13 | US-10-027-632-274762 | Sequence 274762, A |
| 108 | 16.6 | 66.4 | 765     | 16 | US-10-027-632-323670 | Sequence 323670, A | 16 | 181 | 16 | 64.0 | 567  | 16 | US-10-027-632-274762 | Sequence 274762, A |
| 109 | 16.6 | 66.4 | 835     | 13 | US-10-027-632-152146 | Sequence 152146, A | 16 | 182 | 16 | 64.0 | 567  | 16 | US-10-027-632-274762 | Sequence 274762, A |
| 110 | 16.6 | 66.4 | 835     | 16 | US-10-027-632-152146 | Sequence 152146, A | 16 | 183 | 16 | 64.0 | 584  | 15 | US-10-001-883-49     | Sequence 49, Appl  |
| 111 | 16.6 | 66.4 | 835     | 16 | US-10-027-632-152146 | Sequence 152146, A | 16 | 184 | 16 | 64.0 | 584  | 15 | US-09-864-761-8182   | Sequence 8182, Ap  |
| 112 | 16.6 | 66.4 | 1129    | 16 | US-10-369-493-26822  | Sequence 26822, A  | 16 | 185 | 16 | 64.0 | 584  | 9  | US-09-864-761-13596  | Sequence 13596, A  |
| 113 | 16.6 | 66.4 | 1129    | 16 | US-10-261-078-20     | Sequence 20, Appl  | 16 | 186 | 16 | 64.0 | 590  | 15 | US-10-029-386-8674   | Sequence 8674, Ap  |
| 114 | 16.6 | 66.4 | 11987   | 15 | US-08-764-891-9972   | Sequence 9972, Ap  | 16 | 187 | 16 | 64.0 | 590  | 15 | US-10-029-386-8674   | Sequence 8674, Ap  |
| 115 | 16.6 | 66.4 | 12575   | 10 | US-08-764-891-9971   | Sequence 9971, Ap  | 16 | 188 | 16 | 64.0 | 596  | 15 | US-10-029-386-10951  | Sequence 10951, A  |
| 116 | 16.6 | 66.4 | 43793   | 12 | US-09-997-722-37     | Sequence 37, Appl  | 16 | 189 | 16 | 64.0 | 607  | 9  | US-09-809-545A-19    | Sequence 19, Appl  |
| 117 | 16.6 | 66.4 | 256493  | 13 | US-10-087-192-1000   | Sequence 1000, Ap  | 16 | 190 | 16 | 64.0 | 608  | 13 | US-10-027-632-71049  | Sequence 71049, A  |
| 118 | 16.6 | 66.4 | 3186778 | 13 | US-10-027-632-174961 | Sequence 174961, A | 16 | 191 | 16 | 64.0 | 608  | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 119 | 16.6 | 66.4 | 3186778 | 16 | US-10-027-632-174961 | Sequence 174961, A | 16 | 192 | 16 | 64.0 | 608  | 16 | US-10-027-632-71049  | Sequence 71049, A  |
| 120 | 16.4 | 65.6 | 261     | 11 | US-09-864-408A-2621  | Sequence 2621, Ap  | 16 | 193 | 16 | 64.0 | 609  | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 121 | 16.4 | 65.6 | 1494    | 14 | US-10-119-635-3      | Sequence 3, Appl   | 16 | 194 | 16 | 64.0 | 609  | 13 | US-10-027-632-312331 | Sequence 312331, A |
| 122 | 16.4 | 65.6 | 1350    | 14 | US-10-016-157A-126   | Sequence 126, App  | 16 | 195 | 16 | 64.0 | 609  | 16 | US-10-027-632-135529 | Sequence 135529, A |
| 123 | 16.4 | 65.6 | 2240    | 13 | US-10-119-635-7      | Sequence 7, Appl   | 16 | 196 | 16 | 64.0 | 617  | 13 | US-10-027-632-262092 | Sequence 262092, A |
| 124 | 16.4 | 65.6 | 2240    | 14 | US-10-213-796-19     | Sequence 19, Appl  | 16 | 197 | 16 | 64.0 | 617  | 13 | US-10-027-632-262092 | Sequence 262092, A |
| 125 | 16.4 | 65.6 | 51664   | 13 | US-10-087-192-877    | Sequence 877, App  | 16 | 198 | 16 | 64.0 | 617  | 16 | US-10-027-632-262092 | Sequence 262092, A |
| 126 | 16.4 | 65.6 | 94672   | 13 | US-10-087-192-1069   | Sequence 1069, Ap  | 16 | 199 | 16 | 64.0 | 621  | 13 | US-10-282-122A-26617 | Sequence 26617, A  |
| 127 | 16.4 | 65.6 | 106315  | 13 | US-10-087-192-178    | Sequence 178, App  | 16 | 200 | 16 | 64.0 | 621  | 13 | US-10-282-122A-26617 | Sequence 26617, A  |
| 128 | 16.2 | 64.8 | 186     | 15 | US-10-029-386-14920  | Sequence 14920, A  | 16 | 201 | 16 | 64.0 | 624  | 16 | US-10-369-493-31482  | Sequence 31482, A  |
| 129 | 16.2 | 64.8 | 433     | 13 | US-10-027-632-181988 | Sequence 181988, A | 16 | 202 | 16 | 64.0 | 625  | 13 | US-10-027-632-223115 | Sequence 223115, A |
| 130 | 16.2 | 64.8 | 433     | 16 | US-10-027-632-181988 | Sequence 181988, A | 16 | 203 | 16 | 64.0 | 625  | 16 | US-10-027-632-223115 | Sequence 223115, A |
| 131 | 16.2 | 64.8 | 473     | 12 | US-09-876-143-427    | Sequence 427, App  | 16 | 204 | 16 | 64.0 | 634  | 13 | US-10-027-632-134639 | Sequence 134639, A |
| 132 | 16.2 | 64.8 | 542     | 13 | US-10-029-386-1218   | Sequence 1218, Ap  | 16 | 205 | 16 | 64.0 | 634  | 16 | US-10-027-632-134639 | Sequence 134639, A |
| 133 | 16.2 | 64.8 | 542     | 13 | US-10-424-599-83481  | Sequence 83481, A  | 16 | 206 | 16 | 64.0 | 662  | 13 | US-10-027-632-48022  | Sequence 48022, A  |
| 134 | 16.2 | 64.8 | 892     | 13 | US-10-027-632-155735 | Sequence 155735, A | 16 | 207 | 16 | 64.0 | 662  | 16 | US-10-027-632-48022  | Sequence 48022, A  |
| 135 | 16.2 | 64.8 | 892     | 16 | US-10-027-632-155735 | Sequence 155735, A | 16 | 208 | 16 | 64.0 | 689  | 10 | US-09-791-279-20     | Sequence 20, Appl  |
| 136 | 16.2 | 64.8 | 898     | 13 | US-10-027-632-120372 | Sequence 120372, A | 16 | 209 | 16 | 64.0 | 691  | 15 | US-10-029-386-22763  | Sequence 22763, A  |
| 137 | 16.2 | 64.8 | 898     | 16 | US-10-027-632-120372 | Sequence 120372, A | 16 | 210 | 16 | 64.0 | 730  | 13 | US-10-029-386-22763  | Sequence 22763, A  |
| 138 | 16.2 | 64.8 | 931     | 16 | US-10-050-704-88     | Sequence 88, Appl  | 16 | 211 | 16 | 64.0 | 730  | 16 | US-10-027-632-137926 | Sequence 137926, A |
| 139 | 16.2 | 64.8 | 940     | 15 | US-10-050-704-48     | Sequence 48, Appl  | 16 | 212 | 16 | 64.0 | 733  | 13 | US-10-027-632-17850  | Sequence 17850, A  |
| 140 | 16.2 | 64.8 | 1262    | 10 | US-09-984-271-44     | Sequence 44, Appl  | 16 | 213 | 16 | 64.0 | 733  | 16 | US-10-027-632-17850  | Sequence 17850, A  |
| 141 | 16.2 | 64.8 | 1262    | 13 | US-09-984-276-44     | Sequence 44, Appl  | 16 | 214 | 16 | 64.0 | 747  | 13 | US-10-027-632-36534  | Sequence 36534, A  |
| 142 | 16.2 | 64.8 | 1262    | 13 | US-09-984-276-44     | Sequence 44, Appl  | 16 | 215 | 16 | 64.0 | 747  | 16 | US-10-027-632-36534  | Sequence 36534, A  |
| 143 | 16.2 | 64.8 | 2345    | 15 | US-10-106-598-228    | Sequence 228, App  | 16 | 216 | 16 | 64.0 | 925  | 15 | US-10-017-161-1503   | Sequence 1503, Ap  |
| 144 | 16.2 | 64.8 | 4171    | 10 | US-08-764-891-9043   | Sequence 9043, Ap  | 16 | 217 | 16 | 64.0 | 925  | 15 | US-10-017-161-1503   | Sequence 1503, Ap  |
| 145 | 16.2 | 64.8 | 5033    | 16 | US-10-369-300-14     | Sequence 14, Appl  | 16 | 218 | 16 | 64.0 | 1149 | 13 | US-10-424-589-16392  | Sequence 16392, A  |
| 146 | 16.2 | 64.8 | 5610    | 16 | US-10-157-031-127    | Sequence 127, App  | 16 | 219 | 16 | 64.0 | 1149 | 10 | US-09-764-891-7777   | Sequence 7777, Ap  |
| 147 | 16.2 | 64.8 | 47309   | 16 | US-10-085-117-79     | Sequence 79, Appl  | 16 | 220 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7777   | Sequence 7777, Ap  |
| 148 | 16.2 | 64.8 | 66972   | 13 | US-10-087-192-556    | Sequence 556, App  | 16 | 221 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7905   | Sequence 7905, Ap  |
| 149 | 16.2 | 64.8 | 146547  | 15 | US-10-017-128-1      | Sequence 1, Appl   | 16 | 222 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7905   | Sequence 7905, Ap  |
| 150 | 16.2 | 64.8 | 168821  | 13 | US-10-087-192-622    | Sequence 622, App  | 16 | 223 | 16 | 64.0 | 1197 | 10 | US-09-764-891-7909   | Sequence 7909, Ap  |
| 151 | 16.2 | 64.8 | 344548  | 13 | US-10-087-192-334    | Sequence 334, App  | 16 | 224 | 16 | 64.0 | 1278 | 15 | US-10-087-268-6      | Sequence 3, Appl   |
| 152 | 16.2 | 64.8 | 172     | 17 | US-10-712-672-2616   | Sequence 2616, App | 16 | 225 | 16 | 64.0 | 1278 | 15 | US-10-087-268-6      | Sequence 6, Appl   |
| 153 | 16.2 | 64.8 | 172     | 13 | US-10-085-783A-5188  | Sequence 5188, Ap  | 16 | 226 | 16 | 64.0 | 1348 | 17 | US-10-343-710-4      | Sequence 14, Appl  |
| 154 | 16.2 | 64.8 | 172     | 16 | US-10-242-535A-5188  | Sequence 5188, Ap  | 16 | 227 | 16 | 64.0 | 1400 | 17 | US-10-330-051A-36    | Sequence 36, Appl  |
| 155 | 16.2 | 64.8 | 201     | 9  | US-09-983-965-5267   | Sequence 5267, A   | 16 | 228 | 16 | 64.0 | 1497 | 17 | US-10-343-710-35     | Sequence 35, Appl  |
| 156 | 16.2 | 64.8 | 230     | 13 | US-10-085-783A-13031 | Sequence 13031, A  | 16 | 229 | 16 | 64.0 | 1544 | 9  | US-09-349-015-18     | Sequence 18, Appl  |
| 157 | 16.2 | 64.8 | 230     | 16 | US-10-242-535A-13031 | Sequence 13031, A  | 16 | 230 | 16 | 64.0 | 1544 | 15 | US-10-219-664-14     | Sequence 14, Appl  |
| 158 | 16.2 | 64.8 | 294     | 9  | US-09-864-761-19936  | Sequence 19936, A  | 16 | 231 | 16 | 64.0 | 1544 | 15 | US-10-219-664-14     | Sequence 14, Appl  |
| 159 | 16.2 | 64.8 | 321     | 16 | US-10-330-051A-30    | Sequence 30, Appl  | 16 | 232 | 16 | 64.0 | 1804 | 13 | US-10-027-632-100159 | Sequence 1205, Ap  |
| 160 | 16.2 | 64.8 | 323     | 13 | US-10-085-783A-6708  | Sequence 6708, Ap  | 16 | 233 | 16 | 64.0 | 1804 | 16 | US-10-027-632-100159 | Sequence 100159, A |

|       |      |      |        |    |                      |                      |
|-------|------|------|--------|----|----------------------|----------------------|
| C 234 | 16   | 64.0 | 1836   | 13 | US-10-282-122A-12610 | Sequence 12610, A    |
| C 235 | 16   | 64.0 | 1936   | 13 | US-10-296-115-448    | Sequence 448, App    |
| C 236 | 16   | 64.0 | 1949   | 17 | US-10-343-710-15     | Sequence 15, Appl    |
| C 237 | 16   | 64.0 | 2035   | 16 | US-10-094-749-659    | Sequence 659, App    |
| C 238 | 16   | 64.0 | 2044   | 13 | US-10-027-632-97355  | Sequence 97355, A    |
| C 239 | 16   | 64.0 | 2044   | 13 | US-10-027-632-97355  | Sequence 97355, A    |
| C 240 | 16   | 64.0 | 2087   | 17 | US-10-468-334-11     | Sequence 11, Appl    |
| C 241 | 16   | 64.0 | 2174   | 15 | US-10-198-846-10088  | Sequence 10088, A    |
| C 242 | 16   | 64.0 | 2344   | 9  | US-09-880-107-3291   | Sequence 3291, App   |
| C 243 | 16   | 64.0 | 2418   | 9  | US-09-938-842A-872   | Sequence 872, App    |
| C 244 | 16   | 64.0 | 2418   | 11 | US-09-938-842A-872   | Sequence 872, App    |
| C 245 | 16   | 64.0 | 2487   | 9  | US-09-799-514-3      | Sequence 3, Appl     |
| C 246 | 16   | 64.0 | 2534   | 14 | US-10-173-519-1      | Sequence 1, Appl     |
| C 247 | 16   | 64.0 | 2542   | 9  | US-09-964-824A-255   | Sequence 255, App    |
| C 248 | 16   | 64.0 | 2542   | 9  | US-09-954-531-961    | Sequence 961, App    |
| C 249 | 16   | 64.0 | 2681   | 13 | US-10-027-632-112241 | Sequence 112241, App |
| C 250 | 16   | 64.0 | 2681   | 16 | US-10-027-632-112241 | Sequence 112241, App |
| C 251 | 16   | 64.0 | 3147   | 13 | US-10-403-161-71     | Sequence 71, Appl    |
| C 252 | 16   | 64.0 | 3189   | 15 | US-10-017-161-1683   | Sequence 1683, App   |
| C 253 | 16   | 64.0 | 3189   | 16 | US-10-369-493-43549  | Sequence 43549, A    |
| C 254 | 16   | 64.0 | 3225   | 13 | US-10-027-632-115939 | Sequence 115939, App |
| C 255 | 16   | 64.0 | 3225   | 13 | US-10-027-632-115940 | Sequence 115940, App |
| C 256 | 16   | 64.0 | 3225   | 13 | US-10-027-632-115941 | Sequence 115941, App |
| C 257 | 16   | 64.0 | 3225   | 16 | US-10-027-632-115939 | Sequence 115939, App |
| C 258 | 16   | 64.0 | 3225   | 16 | US-10-027-632-115940 | Sequence 115940, App |
| C 259 | 16   | 64.0 | 3225   | 16 | US-10-027-632-115941 | Sequence 115941, App |
| C 260 | 16   | 64.0 | 3584   | 14 | US-10-044-090-341    | Sequence 341, App    |
| C 261 | 16   | 64.0 | 3600   | 15 | US-10-320-786-106    | Sequence 106, App    |
| C 262 | 16   | 64.0 | 3612   | 9  | US-09-890-107-2353   | Sequence 2353, App   |
| C 263 | 16   | 64.0 | 3612   | 16 | US-10-372-683-35     | Sequence 35, Appl    |
| C 264 | 16   | 64.0 | 5373   | 13 | US-10-333-314-40     | Sequence 40, Appl    |
| C 265 | 16   | 64.0 | 5438   | 17 | US-10-601-807-3      | Sequence 3, Appl     |
| C 266 | 16   | 64.0 | 10085  | 10 | US-09-764-891-5491   | Sequence 5491, App   |
| C 267 | 16   | 64.0 | 10136  | 13 | US-10-016-248-1      | Sequence 1, Appl     |
| C 268 | 16   | 64.0 | 23536  | 13 | US-10-087-192-775    | Sequence 775, App    |
| C 269 | 16   | 64.0 | 32607  | 13 | US-10-087-192-106    | Sequence 106, App    |
| C 270 | 16   | 64.0 | 32816  | 9  | US-09-729-094-3      | Sequence 3, Appl     |
| C 271 | 16   | 64.0 | 32816  | 15 | US-10-435-631-3      | Sequence 3, Appl     |
| C 272 | 16   | 64.0 | 70215  | 13 | US-10-087-192-217    | Sequence 217, App    |
| C 273 | 16   | 64.0 | 96588  | 16 | US-10-085-117-76     | Sequence 76, Appl    |
| C 274 | 16   | 64.0 | 134292 | 13 | US-10-240-425-1102   | Sequence 1102, App   |
| C 275 | 16   | 64.0 | 141463 | 13 | US-10-087-192-22     | Sequence 22, Appl    |
| C 276 | 16   | 64.0 | 165961 | 16 | US-10-085-117-241    | Sequence 241, App    |
| C 277 | 16   | 64.0 | 197775 | 13 | US-10-087-192-853    | Sequence 853, App    |
| C 278 | 16   | 64.0 | 220224 | 13 | US-10-087-192-1282   | Sequence 1282, App   |
| C 279 | 16   | 64.0 | 235070 | 13 | US-10-087-192-1990   | Sequence 1990, App   |
| C 280 | 15.8 | 63.2 | 296    | 9  | US-09-933-797-423    | Sequence 423, App    |
| C 281 | 15.8 | 63.2 | 370    | 13 | US-10-424-599-81174  | Sequence 81174, A    |
| C 282 | 15.8 | 63.2 | 421    | 13 | US-10-276-774-58     | Sequence 58, Appl    |
| C 283 | 15.8 | 63.2 | 426    | 9  | US-09-764-877-3239   | Sequence 3239, App   |
| C 284 | 15.8 | 63.2 | 426    | 16 | US-10-242-515-3239   | Sequence 3239, App   |
| C 285 | 15.8 | 63.2 | 448    | 13 | US-10-085-783A-16851 | Sequence 16851, A    |
| C 286 | 15.8 | 63.2 | 448    | 16 | US-10-242-535A-16851 | Sequence 16851, A    |
| C 287 | 15.8 | 63.2 | 480    | 13 | US-10-027-632-289524 | Sequence 289524, App |
| C 288 | 15.8 | 63.2 | 480    | 16 | US-10-027-632-289524 | Sequence 289524, App |
| C 289 | 15.8 | 63.2 | 500    | 14 | US-10-060-841-17     | Sequence 17, Appl    |
| C 290 | 15.8 | 63.2 | 525    | 13 | US-10-027-632-285723 | Sequence 285723, App |
| C 291 | 15.8 | 63.2 | 525    | 16 | US-10-027-632-285723 | Sequence 285723, App |
| C 292 | 15.8 | 63.2 | 563    | 13 | US-10-029-386-13129  | Sequence 13129, A    |
| C 293 | 15.8 | 63.2 | 614    | 13 | US-10-027-632-103268 | Sequence 103268, App |
| C 294 | 15.8 | 63.2 | 614    | 16 | US-10-027-632-103268 | Sequence 103268, App |
| C 295 | 15.8 | 63.2 | 665    | 13 | US-10-027-632-16502  | Sequence 16502, A    |
| C 296 | 15.8 | 63.2 | 665    | 16 | US-10-027-632-16502  | Sequence 16502, A    |
| C 297 | 15.8 | 63.2 | 673    | 13 | US-10-027-632-132497 | Sequence 132497, App |
| C 298 | 15.8 | 63.2 | 673    | 16 | US-10-027-632-132497 | Sequence 132497, App |
| C 299 | 15.8 | 63.2 | 748    | 13 | US-10-027-632-13459  | Sequence 13459, A    |
| C 300 | 15.8 | 63.2 | 748    | 16 | US-10-027-632-13459  | Sequence 13459, A    |

ALIGNMENTS

US-09-712-363-10/c  
; Sequence 10, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-10

Query Match 100.0%; Score 25; DB 9; Length 1584;  
Best Local Similarity 100.0%; Pred. No. 0.11; Indels 0; Gaps 0;  
Matches 25; Conservative 0; Mismatches 0;

QY 1 CCGGAGGGCCAGGGCAGCATAGC 25  
|||  
Db 1019 CCGGAGGGCCAGGGCAGCATAGC 995

RESULT 2  
US-10-271-416-9/c  
; Sequence 9, Application US/10271416  
; Publication No. US20040043021A1  
; GENERAL INFORMATION:  
; APPLICANT: Keith, Tim  
; APPLICANT: Little, Randall D.  
; APPLICANT: Van Erdewegh, Paul  
; APPLICANT: Dupuis, Josee  
; APPLICANT: Del Mastro, Richard G.  
; APPLICANT: Allen, Kristina  
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO RESPIRATORY DISEASES AND OBESITY  
; FILE REFERENCE: 2976-4045  
; CURRENT APPLICATION NUMBER: US/10/271,416  
; CURRENT FILING DATE: 2002-10-11  
; PRIOR APPLICATION NUMBER: 60/328,424  
; PRIOR FILING DATE: 2001-10-11  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 276820  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: conflict

```
; LOCATION: (167043)...(167043)
; OTHER INFORMATION: Y=C or T
US-10-271-416-9

Query Match
Best Local Similarity 80.8%; Score 20.2; DB 13; Length 276820;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 CCGGAGGGCCAGAGGCACATAGC 25
Db 33928 CCGGATGCCAGCGGCGCACAGC 33904

RESULT 3
US-10-027-632-144075
; Sequence 144075, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144075
; LENGTH: 813
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-144075

Query Match
Best Local Similarity 75.2%; Score 18.8; DB 16; Length 813;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 CCGAGGGCCAGAGGCACATA 23
Db 477 CCGAGGGCCAGAGGCACAAA 498

RESULT 5
US-10-052-482-190
; Sequence 190, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 190
; LENGTH: 59817
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (8546)..(8720)
; OTHER INFORMATION: "n" at positions 8546 to 8720 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (18044)..(18090)
; OTHER INFORMATION: "n" at positions 18044 to 18090 can be any base
US-10-052-482-190

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 12; Length 59817;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 CCGGAGGGCCAGAGGCACATAGC 25
Db 8918 CCGGAGGGCCAGAGGCAGCGCGGTAGC 8942

RESULT 6
US-09-960-706-1043/c
; Sequence 1043, Application US/09960706
; Publication No. US20030134280A1
; GENERAL INFORMATION:
; APPLICANT: Murgar, William E.
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic Hyperplasia
; FILE REFERENCE: 44921-5029-01US
```



; CURRENT APPLICATION NUMBER: US/09/960,706  
; CURRENT FILING DATE: 2001-09-24  
; PRIOR APPLICATION NUMBER: 60/223,323  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: 09/873,319  
; PRIOR FILING DATE: 2001-06-05  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1043  
; LENGTH: 14646  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134280A1 X03100  
; NAME/KEY: unsure  
; LOCATION: (1)..(14646)  
; OTHER INFORMATION: n = a or c or g or t  
US-09-960-706-1043

Query Match 73.6%; Score 18.4; DB 10; Length 14646;  
Best Local Similarity 95.0%; Pred. No. 55;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GAGGCCAGAGGGCAGCATAG 24  
Db 2885 GAGGCCAGAGGGACATAG 2866

RESULT 7  
US-09-873-319-691/c  
; Sequence 691, Application US/09873319A  
; Publication No. US20030134324A1  
; GENERAL INFORMATION:  
; APPLICANT: Munger, William E.  
; APPLICANT: Kulkarni, Prakash  
; APPLICANT: Getzenberg, Robert H.  
; APPLICANT: Waga, Iwao  
; APPLICANT: Yamamoto, Jun  
; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
; TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles  
; FILE REFERENCE: 44921-5029-US  
; CURRENT APPLICATION NUMBER: US/09/873,319A  
; CURRENT FILING DATE: 2001-06-05  
; EARLIER APPLICATION NUMBER: US 60/223,323  
; EARLIER FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 755  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 691  
; LENGTH: 14646  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20030134324A1 X03100  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(14646)  
; OTHER INFORMATION: n = a or c or g or t  
US-09-873-319-691

Query Match 73.6%; Score 18.4; DB 10; Length 14646;  
Best Local Similarity 95.0%; Pred. No. 55;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GAGGCCAGAGGGCAGCATAG 24  
Db 2885 GAGGCCAGAGGGACATAG 2866

RESULT 8  
US-10-105-637-1  
; Sequence 1, Application US/10105637  
; Publication No. US2003008752A1  
; GENERAL INFORMATION:

; APPLICANT: Morris, David  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER  
; TITLE OF INVENTION: ASSOCIATED WITH ALTERED EXPRESSION OF PRDM11  
; FILE REFERENCE: 529452500120  
; CURRENT APPLICATION NUMBER: US/10/105,637  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: US 10/034,650  
; PRIOR FILING DATE: 2001-12-20  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 90442  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-105-637-1

Query Match 72.8%; Score 18.2; DB 15; Length 90442;  
Best Local Similarity 87.0%; Pred. No. 49;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCGAGGGCCAGAGGGCAGCATAG 24  
Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 9  
US-10-034-650-43  
; Sequence 43, Application US/10034650  
; Publication No. US20030216558A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David  
; APPLICANT: Engelhard, Eric  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000128  
; CURRENT APPLICATION NUMBER: US/10/034,650  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: US 09/474,377  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 43  
; LENGTH: 90442  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-034-650-43

Query Match 72.8%; Score 18.2; DB 16; Length 90442;  
Best Local Similarity 87.0%; Pred. No. 49;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CCGAGGGCCAGAGGGCAGCATAG 24  
Db 27716 CCAGAGGGACAGAGGGCAGATAG 27738

RESULT 10  
US-10-029-386-11131/c  
; Sequence 11131, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

FILE REFERENCE: ABOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 11131  
LENGTH: 523  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004178.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.51  
OTHER INFORMATION: SWISSPROT HIT: P35410, EVALUE 5.00e-23  
OTHER INFORMATION: EST HUMAN HIT: BF883760.1, EVALUE 7.30e-01  
OTHER INFORMATION: NT HIT: S78653.1, EVALUE 3.20e-02  
US-10-029-386-11131

Query Match 70.4%; Score 17.6; DB 15; Length 523;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGGCCACATAG 24  
Db 413 CCAGGTGGACAGAGGGCCACAGAG 390

RESULT 11  
US-09-791-279-49  
Sequence 49, Application US/09791279  
Publication No. US20030050456A1  
GENERAL INFORMATION:  
APPLICANT: Vogeli, Gabriel  
APPLICANT: Wood, Linda S.  
APPLICANT: Parodi Luis  
APPLICANT: Lind, Peter  
TITLE OF INVENTION: No. US20030050456A1 G Protein-Coupled Receptors  
FILE REFERENCE: 00048.US1  
CURRENT APPLICATION NUMBER: US/09/791,279  
CURRENT FILING DATE: 2001-02-23  
PRIOR APPLICATION NUMBER: 60/184,715  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,725  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,712  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,606  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,602  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,604  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,822  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,710  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,689  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,690  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: 60/184,716  
PRIOR FILING DATE: 2000-02-24  
NUMBER OF SEQ ID NOS: 220  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 49  
LENGTH: 637  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-791-279-49

Query Match 70.4%; Score 17.6; DB 10; Length 637;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGGCCACATAG 24  
Db 182 CCAGGTGGACAGAGGGCCACAGAG 205  
RESULT 12  
US-10-027-632-140370/C  
Sequence 140370, Application US/10027632  
Publication No. US20020198371A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 140370  
LENGTH: 834  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-140370

Query Match 70.4%; Score 17.6; DB 13; Length 834;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGGCCACATAG 24  
Db 154 CCAGAGGGACAGAGGGCCACATGG 131  
RESULT 13  
US-10-027-632-140370/C  
Sequence 140370, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720

Query Match 70.4%; Score 17.6; DB 13; Length 834;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGGCCACATAG 24  
Db 154 CCAGAGGGACAGAGGGCCACATGG 131  
RESULT 13  
US-10-027-632-140370/C  
Sequence 140370, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720

Query Match 70.4%; Score 17.6; DB 13; Length 834;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGGCCACATAG 24  
Db 154 CCAGAGGGACAGAGGGCCACATGG 131  
RESULT 13  
US-10-027-632-140370/C  
Sequence 140370, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720

Query Match 70.4%; Score 17.6; DB 10; Length 637;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGGCCACATAG 24  
Db 154 CCAGAGGGACAGAGGGCCACATGG 131  
RESULT 13  
US-10-027-632-140370/C  
Sequence 140370, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720

Query Match 70.4%; Score 17.6; DB 10; Length 637;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCCGAGGGCCACAGAGGGCCACATAG 24  
Db 154 CCAGAGGGACAGAGGGCCACATGG 131  
RESULT 13  
US-10-027-632-140370/C  
Sequence 140370, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720

Query Match 70.4%; Score 17.6; DB 10; Length 637;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 140370
; LENGTH: 834
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140370

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 16; Length 834;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACATAG 24
Db 154 CCCAGAGGACAGAGGGCAGATGG 131

RESULT 14
US-10-027-632-251702
; Sequence 251702, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 251702
; LENGTH: 1010
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-251702

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 16; Length 1010;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACATAG 24
Db 380 CCCGAGAGGCGAGGGGCACACAG 403

RESULT 16
US-10-017-161-1679/c
; Sequence 1679, Application US/10017161
; Publication No. US20030143688A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABEURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1679
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1152)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(592)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (781)..(952)
US-10-017-161-1679

Query Match
Best Local Similarity 70.4%; Score 17.6; DB 15; Length 1152;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCACATAG 24
Db 554 CCAGGTGGGACAGAGGGGCACAGAG 531

RESULT 17
US-10-292-798-1339/c
; Sequence 1339, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
```

APPLICANT: SUWA, MAKIKO  
APPLICANT: ASAI, KIYOSHI  
APPLICANT: AKIYAMA, YUTAKA  
APPLICANT: ABURATANI, HIROYUKI  
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
FILE REFERENCE: 084335/166  
CURRENT APPLICATION NUMBER: US/10/292,798  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: 10/017,161  
PRIOR FILING DATE: 2001-12-18  
PRIOR APPLICATION NUMBER: JP 2001-246789  
PRIOR FILING DATE: 2001-06-18  
NUMBER OF SEQ ID NOS: 2070  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1339  
LENGTH: 1152  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
LOCATION: source  
FEATURE:  
LOCATION: (1)..(1152)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (201)..(629)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (728)..(952)  
US-10-292-798-1339

Query Match 70.4%; Score 17.6; DB 16; Length 1152;  
Best Local Similarity 83.3%; Pred. No. 1.9e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCACAGAGGCACATAG 24  
DB 554 CCAGGTGGACAGAGGGCCACAGAG 531

RESULT 18  
US-10-632-112152  
Sequence 112152, Application US/10027632  
Publication No. US20020198371A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 112152  
LENGTH: 2545  
TYPE: DNA  
ORGANISM: Human  
US-10-632-112152

Query Match 70.4%; Score 17.6; DB 13; Length 2545;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCACAGAGGCACATAG 24  
DB 907 CCCAGAGGGACAGAGGGCCAGATGG 930

RESULT 19  
US-10-632-112153  
Sequence 112153, Application US/10027632  
Publication No. US20020198371A1  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 112153  
LENGTH: 2545  
TYPE: DNA  
ORGANISM: Human  
US-10-632-112153

Query Match 70.4%; Score 17.6; DB 13; Length 2545;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGAGGGCCACAGAGGCACATAG 24  
DB 907 CCCAGAGGGACAGAGGGCCAGATGG 930

RESULT 20  
US-10-632-112152  
Sequence 112152, Application US/10027632  
Publication No. US20020204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 112152  
; LENGTH: 2545  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-112152

Query Match 70.4%; Score 17.6; DB 16; Length 2545;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCCACATAG 24  
Db 907 CCAGAGGGACAGAGGGCCACATGG 930

RESULT 21  
US-10-027-632-112153  
; Sequence 112153, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027.632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 112153  
; LENGTH: 2545  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-112153

Query Match 70.4%; Score 17.6; DB 16; Length 2545;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CCGGAGGGCCAGAGGGCCACATAG 24  
Db 907 CCAGAGGGACAGAGGGCCACATGG 930

RESULT 22  
US-10-087-192-1045  
; Sequence 1045, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087.192  
; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1045  
; LENGTH: 24902  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...((24902))  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-1045

Query Match 69.6%; Score 17.4; DB 13; Length 24902;  
Best Local Similarity 94.7%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 AGGGCCAGAGGGCCACATAG 24  
Db 504 AGGGCCAGAGGGCCACAGAG 522

RESULT 23  
US-10-424-599-139706  
; Sequence 139706, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 139706  
; LENGTH: 440  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_97162C.1  
US-10-424-599-139706

Query Match 68.8%; Score 17.2; DB 13; Length 440;  
Best Local Similarity 86.4%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CCGAGGGCCAGAGGGCCACATAG 24  
Db 109 CCGAGGGCCCTGTGAGGCACATAG 130

RESULT 24  
US-10-027-632-144076/c  
; Sequence 144076, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108927.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 144076  
; LENGTH: 733  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-144076

Query Match 68.8%; Score 17.2; DB 13; Length 733;  
Best Local Similarity 86.4%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCACATA 23  
DB 515 CAGGAAGGCCAGAGGGCACAAA 494

RESULT 25  
US-10-027-632-144076/c  
; Sequence 144076, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827,129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 144076  
; LENGTH: 733  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-144076

Query Match 68.8%; Score 17.2; DB 16; Length 733;  
Best Local Similarity 86.4%; Pred. No. 3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCACATA 23  
DB 515 CAGGAAGGCCAGAGGGCACAAA 494

RESULT 26  
US-10-429-223-1  
; Sequence 1, Application US/10429223  
; Publication No. US20040009513A1  
; GENERAL INFORMATION:

; APPLICANT: Vatner, Stephen  
; APPLICANT: Depre, Christophe  
; TITLE OF INVENTION: pDJAI, a cardiac specific gene,  
; TITLE OF INVENTION: corresponding proteins and uses thereof  
; FILE REFERENCE: 601-1-130  
; CURRENT APPLICATION NUMBER: US/10/429,223  
; PRIOR FILING DATE: 2003-05-02  
; PRIOR APPLICATION NUMBER: 60/377,578  
; PRIOR FILING DATE: 2002-05-02  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 3014  
; TYPE: DNA  
; ORGANISM: Sus scrofa  
; NAME/KEY: misc\_feature  
; LOCATION: 16  
; OTHER INFORMATION: n = A,T,C or G  
US-10-429-223-1

Query Match 68.8%; Score 17.2; DB 16; Length 3014;  
Best Local Similarity 86.4%; Pred. No. 2.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCGAGGGCCAGAGGGCACAT 22  
DB 2592 CCGAGGGCCAGAGGGCACAT 2613

RESULT 27  
US-10-052-482-13  
; Sequence 13, Application US/10052482  
; Publication No. US20040072264A1  
; GENERAL INFORMATION:  
; APPLICANT: Engelhard, Eric  
; APPLICANT: Morris, David  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
; FILE REFERENCE: A-71087/RMS/DCF  
; CURRENT APPLICATION NUMBER: US/10/052,482  
; PRIOR FILING DATE: 2002-08-15  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 241  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 13  
; LENGTH: 23070  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; NAME/KEY: misc\_feature  
; LOCATION: (3127)..(3343)  
; OTHER INFORMATION: "n" at positions 3127 to 3343 can be any base.  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (17723)..(17758)  
; OTHER INFORMATION: "n" at positions 17723 to 17758 can be any base.  
US-10-052-482-13

Query Match 68.8%; Score 17.2; DB 12; Length 23070;  
Best Local Similarity 86.4%; Pred. No. 1.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CCGAGGGCCAGAGGGCACATA 23  
DB 18715 CCGAGGGCCAGAGGGCACATA 18736

## RESULT 28

US-09-925-302-409/c  
; Sequence 409, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 409  
; LENGTH: 376  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (223)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (246)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (250)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (259)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (273)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (291)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (361)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (364)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (367)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (372)  
; OTHER INFORMATION: n equals a,t,g, or c  
; US-09-925-302-409

Query Match 68.0%; Score 17; DB 9; Length 376;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GAGGGCCAGAGGGCACA 21  
Db 143 GAGGGCCAGAGGGCACA 127

## RESULT 29

US-09-925-302-409/c  
; Sequence 409, Application US/09925302  
; Patent No. US20030064072A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; CURRENT FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 409  
; LENGTH: 376  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (223)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (246)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (250)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (259)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (273)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (291)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (361)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (364)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (367)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (372)  
; OTHER INFORMATION: n equals a,t,g, or c  
; US-09-925-302-409

Query Match 68.0%; Score 17; DB 13; Length 376;  
Best Local Similarity 100.0%; Pred. No. 4.1e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 GAGGGCCAGAGGGCACA 21  
Db 143 GAGGGCCAGAGGGCACA 127

## RESULT 30

US-09-918-995-6988/c  
; Sequence 6988, Application US/09918995  
; Publication No. US2003007623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6988  
; LENGTH: 406  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-918-995-6988

Query Match 68.0%; Score 17; DB 10; Length 406;  
 Best Local Similarity 100.0%; Pred. No. 4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 GAGGCCACAGAGGCCACA 21  
 |||||  
 Db 73 GAGGCCACAGAGGCCACA 57

Search completed: June 20, 2004, 17:46:38  
 Job time : 151.402 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:50:47 ; Search time 914.603 Seconds  
(without alignments)  
1596.226 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcgagaccaggtagcaccatgag 30

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hug.\*

3: gb\_in.\*

4: gb\_cm.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pi.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pi.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_un.\*

28: em\_vi.\*

29: em\_hug\_hum.\*

30: em\_hug\_inv.\*

31: em\_hug\_other.\*

32: em\_hug\_mus.\*

33: em\_hug\_pln.\*

34: em\_hug\_rod.\*

35: em\_hug\_vrt.\*

36: em\_hug\_mam.\*

37: em\_sy.\*

38: em\_hgo\_hum.\*

39: em\_hgo\_mus.\*

40: em\_hgo\_pln.\*

41: em\_hgo\_rod.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID       | Description          |
|------------|-------|---------------|--------|----|----------|----------------------|
| C 1        | 30    | 100.0         | 15619  | 1  | AS006916 | Mycobacte            |
| C 2        | 30    | 100.0         | 38721  | 1  | MSGY219  | AD000013 Mycobacte   |
| C 3        | 30    | 100.0         | 341957 | 15 | BX842572 | BX842572 Mycobacte   |
| C 4        | 30    | 100.0         | 343050 | 1  | BX248334 | BX248334 Mycobacte   |
| C 5        | 23.8  | 79.3          | 37821  | 1  | MLCB1770 | Z70722 Mycobacteri   |
| C 6        | 23.8  | 79.3          | 344050 | 1  | MLBPRN1  | AL583917 Mycobacte   |
| C 7        | 21.2  | 70.7          | 98512  | 9  | AC022539 | AC022539 Homo sapi   |
| C 8        | 21.2  | 70.7          | 154359 | 2  | AC021178 | AC021178 Homo sapi   |
| C 9        | 21.2  | 70.7          | 158559 | 9  | AC091772 | AC091772 Homo sapi   |
| C 10       | 21.2  | 70.7          | 158568 | 9  | AC069541 | AC069541 Homo sapi   |
| C 11       | 21.2  | 70.7          | 188591 | 2  | AC118678 | AC118678 Mus muscu   |
| C 12       | 21.2  | 70.0          | 188591 | 2  | AC105077 | AC105077 Mus muscu   |
| C 13       | 20.6  | 68.7          | 510    | 6  | AR413807 | AR413807 Sequence    |
| C 14       | 20.6  | 68.7          | 510    | 6  | BD109360 | BD109360 EST and e   |
| C 15       | 20.6  | 68.7          | 1476   | 6  | E10616   | E10616 Human CDNA    |
| C 16       | 20.6  | 68.7          | 1476   | 6  | E10861   | E10861 cDNA encodi   |
| C 17       | 20.6  | 68.7          | 1476   | 6  | AR399329 | AR399329 Sequence    |
| C 18       | 20.6  | 68.7          | 1520   | 9  | AF182277 | AF182277 Homo sapi   |
| C 19       | 20.6  | 68.7          | 2000   | 9  | AF081569 | AF081569 Homo sapi   |
| C 20       | 20.6  | 68.7          | 3045   | 6  | AX018551 | AX018551 Sequence    |
| C 21       | 20.6  | 68.7          | 3045   | 6  | AX018627 | AX018627 Sequence    |
| C 22       | 20.6  | 68.7          | 3045   | 6  | AX332736 | AX332736 Sequence    |
| C 23       | 20.6  | 68.7          | 3045   | 6  | AX409672 | AX409672 Sequence    |
| C 24       | 20.6  | 68.7          | 3045   | 6  | BD136285 | BD136285 Enhanceme   |
| C 25       | 20.6  | 68.7          | 3045   | 6  | HUMCYB2B | M29874 Human cyto    |
| C 26       | 20.6  | 68.7          | 39944  | 9  | AC011541 | AC011541 Homo sapi   |
| C 27       | 20.6  | 68.7          | 42547  | 9  | AC023172 | AC023172 Homo sapi   |
| C 28       | 20.6  | 68.7          | 150962 | 9  | AC092374 | AC092374 Homo sapi   |
| C 29       | 20.6  | 68.7          | 164831 | 2  | AC145155 | AC145155 Homo sapi   |
| C 30       | 20.6  | 68.7          | 222809 | 10 | AC117614 | AC117614 Mus muscu   |
| C 31       | 20.4  | 68.0          | 41     | 6  | AX214506 | AX214506 Sequence    |
| C 32       | 20.4  | 68.0          | 1595   | 4  | RABPEA   | L08418 Cryptotolagus |
| C 33       | 20.4  | 68.0          | 33385  | 2  | AC013058 | AC013058 Drosophil   |
| C 34       | 20.4  | 68.0          | 160294 | 2  | AC025621 | AC025621 Homo sapi   |
| C 35       | 20.4  | 68.0          | 165654 | 2  | AC027747 | AC027747 Homo sapi   |
| C 36       | 20.4  | 68.0          | 165928 | 3  | AC023725 | AC023725 Drosophil   |
| C 37       | 20.4  | 68.0          | 168208 | 2  | AC024359 | AC024359 Homo sapi   |
| C 38       | 20.4  | 68.0          | 172784 | 3  | AC105352 | AC105352 Drosophil   |
| C 39       | 20.4  | 68.0          | 176921 | 9  | AC016080 | AC016080 Homo sapi   |
| C 40       | 20.4  | 68.0          | 178142 | 9  | AC013546 | AC013546 Homo sapi   |
| C 41       | 20.4  | 68.0          | 193572 | 10 | AC140268 | AC140268 Mus muscu   |
| C 42       | 20.4  | 68.0          | 200363 | 9  | AC010895 | AC010895 Homo sapi   |
| C 43       | 20.4  | 68.0          | 214271 | 2  | AC090737 | AC090737 Homo sapi   |
| C 44       | 20.4  | 68.0          | 216834 | 10 | AC123867 | AC123867 Mus muscu   |
| C 45       | 20.4  | 68.0          | 220715 | 9  | AC012410 | AC012410 Homo sapi   |
| C 46       | 20.4  | 68.0          | 230614 | 10 | AL626772 | AL626772 Mouse DNA   |
| C 47       | 20.4  | 68.0          | 233913 | 2  | AC124048 | AC124048 Rattus no   |
| C 48       | 20.4  | 68.0          | 246196 | 2  | AC095869 | AC095869 Rattus no   |
| C 49       | 20.4  | 68.0          | 249044 | 2  | AC103086 | AC103086 Rattus no   |
| C 50       | 20.4  | 68.0          | 295177 | 3  | AE003430 | AE003430 Drosophil   |
| C 51       | 20.2  | 67.3          | 76681  | 9  | AL592223 | AL592223 Human DNA   |
| C 52       | 20.2  | 67.3          | 163183 | 9  | AC016574 | AC016574 Homo sapi   |
| C 53       | 20.2  | 67.3          | 176447 | 2  | AC024894 | AC024894 Homo sapi   |
| C 54       | 20.2  | 67.3          | 179648 | 10 | AL672122 | AL672122 Mouse DNA   |
| C 55       | 20.2  | 67.3          | 203134 | 9  | AC021077 | AC021077 Homo sapi   |
| C 56       | 20.2  | 67.3          | 348050 | 1  | AP003581 | AP003581 Nostoc sp   |
| C 57       | 20    | 66.7          | 1588   | 8  | AY051044 | AY051044 Arabidops   |
| C 58       | 20    | 66.7          | 2374   | 8  | AF360288 | AF360288 Arabidops   |
| C 59       | 20    | 66.7          | 53000  | 2  | AC003656 | Continuation (8 of   |
| C 60       | 20    | 66.7          | 74350  | 8  | AP002543 | Continuation (8 of   |
| C 61       | 20    | 66.7          | 77508  | 2  | AC023005 | Continuation (7 of   |
| C 62       | 20    | 66.7          | 110000 | 2  | AC003656 | Continuation (7 of   |
| C 63       | 20    | 66.7          | 120871 | 9  | AP001054 | AP001054 Homo sapi   |
| C 64       | 20    | 66.7          | 122638 | 9  | AB001523 | AB001523 Homo sapi   |
| C 65       | 20    | 66.7          | 168277 | 2  | AC144598 | AC144598 Ornithorh   |

|   |     |      |      |        |    |           |                    |       |      |      |        |    |           |                     |
|---|-----|------|------|--------|----|-----------|--------------------|-------|------|------|--------|----|-----------|---------------------|
| C | 66  | 20   | 66.7 | 174503 | 9  | AC005697  | AC005697 Homo sapi | C 139 | 19.4 | 64.7 | 188928 | 10 | AC121814  | AC121814 Mus muscu  |
| C | 67  | 20   | 66.7 | 177906 | 2  | AC102047  | AC102047 Mus muscu | C 140 | 19.4 | 64.7 | 190173 | 10 | AC110573  | AC110573 Homo sapi  |
| C | 68  | 20   | 66.7 | 177917 | 2  | AC074236  | AC074236 Homo sapi | C 141 | 19.4 | 64.7 | 191154 | 2  | AC132150  | AC132150 Mus muscu  |
| C | 69  | 20   | 66.7 | 183890 | 9  | AC090868  | AC090868 Homo sapi | C 142 | 19.4 | 64.7 | 193694 | 2  | AC120728  | AC120728 Rattus no  |
| C | 70  | 20   | 66.7 | 188561 | 2  | AC115738  | AC115738 Mus muscu | C 143 | 19.4 | 64.7 | 196784 | 2  | AC113006  | AC113006 Mus muscu  |
| C | 71  | 20   | 66.7 | 194101 | 2  | AC090910  | AC090910 Homo sapi | C 144 | 19.4 | 64.7 | 198093 | 9  | AC018768  | AC018768 Homo sapi  |
| C | 72  | 20   | 66.7 | 220888 | 2  | AC136582  | AC136582 Rattus no | C 145 | 19.4 | 64.7 | 199113 | 2  | AC120411  | AC120411 Mus muscu  |
| C | 73  | 20   | 66.7 | 239292 | 2  | AC108339  | AC108339 Rattus no | C 146 | 19.4 | 64.7 | 200711 | 2  | AC133960  | AC133960 Homo sapi  |
| C | 74  | 20   | 66.7 | 240971 | 2  | AC130253  | AC130253 Rattus no | C 147 | 19.4 | 64.7 | 215863 | 2  | AC136812  | AC136812 Rattus no  |
| C | 75  | 20   | 66.7 | 340000 | 9  | AP001753  | AP001753 Homo sapi | C 148 | 19.4 | 64.7 | 224446 | 2  | AC132531  | AC132531 Rattus no  |
| C | 76  | 19.8 | 66.0 | 3432   | 14 | TACNP     | M20304 Tacaribe vi | C 149 | 19.4 | 64.7 | 227162 | 2  | AC115138  | AC115138 Rattus no  |
| C | 77  | 19.8 | 66.0 | 103019 | 5  | AC091292  | AC091292 Takifugu  | C 150 | 19.4 | 64.7 | 227976 | 2  | AC134029  | AC134029 Rattus no  |
| C | 78  | 19.8 | 66.0 | 215946 | 10 | AC124590  | AC124590 Mus muscu | C 151 | 19.4 | 64.7 | 228242 | 2  | AC109115  | AC109115 Rattus no  |
| C | 79  | 19.6 | 65.3 | 590    | 6  | AX184940  | AX184940 Sequence  | C 152 | 19.4 | 64.7 | 238727 | 2  | AC114539  | AC114539 Mus muscu  |
| C | 80  | 19.6 | 65.3 | 649    | 6  | AX187808  | AX187808 Sequence  | C 153 | 19.4 | 64.7 | 239467 | 2  | AC111384  | AC111384 Rattus no  |
| C | 81  | 19.6 | 65.3 | 695    | 11 | BV048790  | BV048790 S212P6227 | C 154 | 19.4 | 64.7 | 240315 | 2  | AC099474  | AC099474 Rattus no  |
| C | 82  | 19.6 | 65.3 | 778    | 6  | AX186743  | AX186743 Sequence  | C 155 | 19.4 | 64.7 | 246402 | 2  | AC134800  | AC134800 Rattus no  |
| C | 83  | 19.6 | 65.3 | 785    | 10 | AF044280  | AF044280 Rattus no | C 156 | 19.4 | 64.7 | 248439 | 2  | AC098173  | AC098173 Rattus no  |
| C | 84  | 19.6 | 65.3 | 2569   | 9  | AC091287  | AC091287 Homo sapi | C 157 | 19.4 | 64.7 | 248527 | 2  | AC107098  | AC107098 Rattus no  |
| C | 85  | 19.6 | 65.3 | -4772  | 3  | BT003507  | BT003507 Drosophil | C 158 | 19.4 | 64.7 | 249230 | 2  | AC121466  | AC121466 Rattus no  |
| C | 86  | 19.6 | 65.3 | 6248   | 10 | AX122196  | AX122196 Mus muscu | C 159 | 19.4 | 64.7 | 253346 | 2  | AC128440  | AC128440 Rattus no  |
| C | 87  | 19.6 | 65.3 | 6847   | 9  | HS806465  | HS806465 Sequence  | C 160 | 19.4 | 64.7 | 255746 | 2  | AC122622  | AC122622 Rattus no  |
| C | 88  | 19.6 | 65.3 | 11191  | 2  | AC020352  | AC020352 Drosophil | C 161 | 19.4 | 64.7 | 256813 | 2  | AC135301  | AC135301 Rattus no  |
| C | 89  | 19.6 | 65.3 | 111107 | 9  | AC005924  | AC005924 Homo sapi | C 162 | 19.4 | 64.7 | 261538 | 2  | AC112004  | AC112004 Rattus no  |
| C | 90  | 19.6 | 65.3 | 145414 | 9  | HS78F24   | AL022336 Human DNA | C 163 | 19.4 | 64.7 | 272627 | 2  | AC129054  | AC129054 Rattus no  |
| C | 91  | 19.6 | 65.3 | 150587 | 3  | AC007549  | AC007549 Drosophil | C 164 | 19.4 | 64.7 | 316613 | 8  | SCCHR111  | SCCHR111            |
| C | 92  | 19.6 | 65.3 | 159094 | 5  | AL929193  | AL929193 Zebrafish | C 165 | 19.2 | 64.0 | 912    | 8  | AF106929  | AF106929 Medicago   |
| C | 93  | 19.6 | 65.3 | 162522 | 2  | AC021014  | AC021014 Homo sapi | C 166 | 19.2 | 64.0 | 1136   | 8  | D89219    | D89219 Schizosacch  |
| C | 94  | 19.6 | 65.3 | 169816 | 2  | AC118600  | AC118600 Mus muscu | C 167 | 19.2 | 64.0 | 1210   | 8  | TOBFESA   | TOBFESA             |
| C | 95  | 19.6 | 65.3 | 175326 | 9  | CNS01DTQ  | AL132838 Human chr | C 168 | 19.2 | 64.0 | 1286   | 8  | TOBPESC   | TOBPESC             |
| C | 96  | 19.6 | 65.3 | 181771 | 3  | AC008340  | AC008340 Drosophil | C 169 | 19.2 | 64.0 | 1598   | 10 | BC010459  | BC010459 Mus muscu  |
| C | 97  | 19.6 | 65.3 | 198100 | 2  | BX897683  | BX897683 Mus muscu | C 170 | 19.2 | 64.0 | 1655   | 10 | MNNEFA    | MNNEFA              |
| C | 98  | 19.6 | 65.3 | 198772 | 10 | AL606906  | AL606906 Mouse DNA | C 171 | 19.2 | 64.0 | 2250   | 8  | AB012387  | AB012387 Schizosac  |
| C | 99  | 19.6 | 65.3 | 205268 | 2  | BX890581  | BX890581 Danio rer | C 172 | 19.2 | 64.0 | 2870   | 3  | AF029249  | AF029249 Mytilus e  |
| C | 100 | 19.6 | 65.3 | 209197 | 9  | AC007389  | AC007389 Homo sapi | C 173 | 19.2 | 64.0 | 3328   | 8  | AF000150  | AF000150 Schizosac  |
| C | 101 | 19.6 | 65.3 | 221292 | 9  | AF288742  | AF288742 Homo sapi | C 174 | 19.2 | 64.0 | 4039   | 10 | AF081499  | AF081499 Mus muscu  |
| C | 102 | 19.6 | 65.3 | 225548 | 2  | AC099714  | AC099714 Mus muscu | C 175 | 19.2 | 64.0 | 10029  | 1  | AE012085  | AE012085 Xanthomon  |
| C | 103 | 19.6 | 65.3 | 237178 | 2  | AC105297  | AC105297 Mus muscu | C 176 | 19.2 | 64.0 | 13630  | 1  | AE007459  | AE007459 Streptococ |
| C | 104 | 19.6 | 65.3 | 242649 | 1  | AC108648  | AC108648 Rattus no | C 177 | 19.2 | 64.0 | 14089  | 8  | SPAPB18E9 | SPAPB18E9           |
| C | 105 | 19.6 | 65.3 | 300956 | 1  | AE016963  | AE016963 Coccidia  | C 178 | 19.2 | 64.0 | 21706  | 6  | AR218804  | AR218804 Sequence   |
| C | 106 | 19.6 | 65.3 | 348873 | 3  | AE003790  | AE003790 Drosophil | C 179 | 19.2 | 64.0 | 21706  | 6  | BD003716  | BD003716 Polyucle   |
| C | 107 | 19.4 | 64.7 | 655    | 11 | BV018096  | BV018096 S208P6653 | C 180 | 19.2 | 64.0 | 39781  | 8  | SPCC1739  | SPCC1739            |
| C | 108 | 19.4 | 64.7 | 1260   | 14 | TRVAPGD   | L34031 Turkey rhin | C 181 | 19.2 | 64.0 | 110000 | 2  | LMFLCHR32 | LMFLCHR32           |
| C | 109 | 19.4 | 64.7 | 1260   | 14 | TRVAPGD   | L34033 Turkey rhin | C 182 | 19.2 | 64.0 | 158764 | 10 | AC132083  | AC132083 Homo sapi  |
| C | 110 | 19.4 | 64.7 | 1260   | 14 | TRVAPGD   | L34034 Turkey rhin | C 183 | 19.2 | 64.0 | 158952 | 9  | AC096745  | AC096745 Mus muscu  |
| C | 111 | 19.4 | 64.7 | 1485   | 3  | PFACSNURI | M11031 Plasmodium  | C 184 | 19.2 | 64.0 | 172425 | 2  | AC124108  | AC124108 Mus muscu  |
| C | 112 | 19.4 | 64.7 | 1587   | 10 | BC049142  | BC049142 Mus muscu | C 185 | 19.2 | 64.0 | 174363 | 2  | AC144639  | AC144639 Ornithorh  |
| C | 113 | 19.4 | 64.7 | 2261   | 10 | BC058745  | BC058745 Mus muscu | C 186 | 19.2 | 64.0 | 185695 | 2  | AC107837  | AC107837 Mus muscu  |
| C | 114 | 19.4 | 64.7 | 3337   | 8  | AC065496  | AC065496 Oryza sat | C 187 | 19.2 | 64.0 | 210172 | 2  | AC127686  | AC127686 Mus muscu  |
| C | 115 | 19.4 | 64.7 | 3465   | 8  | SCABPI    | XS1780 Yeast ABPI  | C 188 | 19.2 | 64.0 | 248489 | 2  | AC115656  | AC115656 Rattus no  |
| C | 116 | 19.4 | 64.7 | 51420  | 9  | AL139412  | AL139412 Human DNA | C 189 | 19.2 | 64.0 | 249625 | 2  | AC094817  | AC094817 Rattus no  |
| C | 117 | 19.4 | 64.7 | 95845  | 8  | AF411807  | AF411807 Lycopersi | C 190 | 19.2 | 64.0 | 254098 | 2  | AC096236  | AC096236 Rattus no  |
| C | 118 | 19.4 | 64.7 | 96180  | 9  | AC005868  | AC005868 Homo sapi | C 191 | 19.2 | 64.0 | 262405 | 2  | AC094802  | AC094802 Rattus no  |
| C | 119 | 19.4 | 64.7 | 99003  | 2  | AL390756  | AL390756 Homo sapi | C 192 | 19.2 | 64.0 | 266051 | 2  | AC094802  | AC094802 Rattus no  |
| C | 120 | 19.4 | 64.7 | 120770 | 9  | AL513190  | AL513190 Human DNA | C 193 | 19.2 | 64.0 | 274906 | 2  | AC094362  | AC094362 Rattus no  |
| C | 121 | 19.4 | 64.7 | 125750 | 2  | AC092011  | AC092011 Felis cat | C 194 | 19.2 | 64.0 | 349880 | 6  | AX571764  | AX571764 Sequence   |
| C | 122 | 19.4 | 64.7 | 14653  | 10 | AC108919  | AC108919 Mus muscu | C 195 | 19.2 | 63.3 | 280    | 6  | AX904472  | AX904472 Sequence   |
| C | 123 | 19.4 | 64.7 | 151700 | 9  | AC133528  | AC133528 Homo sapi | C 196 | 19.2 | 63.3 | 280    | 6  | BSA299740 | BSA299740 Homo sapi |
| C | 124 | 19.4 | 64.7 | 152299 | 2  | AC125012  | AC125012 Mus muscu | C 197 | 19.2 | 63.3 | 1167   | 9  | BC005050  | BC005050 Homo sapi  |
| C | 125 | 19.4 | 64.7 | 153675 | 2  | AC092012  | AC092012 Felis cat | C 198 | 19.2 | 63.3 | 1578   | 9  | BC005050  | BC005050 Homo sapi  |
| C | 126 | 19.4 | 64.7 | 153675 | 2  | AC139188  | AC139188 Rattus no | C 199 | 19.2 | 63.3 | 2019   | 9  | BC050853  | BC050853 Homo sapi  |
| C | 127 | 19.4 | 64.7 | 153675 | 9  | AC018767  | AC018767 Homo sapi | C 200 | 19.2 | 63.3 | 2128   | 4  | AF074085  | AF074085 Felis cat  |
| C | 128 | 19.4 | 64.7 | 153675 | 10 | AL531075  | AL531075 Mouse DNA | C 201 | 19.2 | 63.3 | 2318   | 9  | AK094248  | AK094248 Homo sapi  |
| C | 129 | 19.4 | 64.7 | 160545 | 2  | AP003508  | AP003508 Oryza sat | C 202 | 19.2 | 63.3 | 2377   | 9  | AF538150  | AF538150 Homo sapi  |
| C | 130 | 19.4 | 64.7 | 160557 | 2  | AC021795  | AC021795 Homo sapi | C 203 | 19.2 | 63.3 | 31004  | 9  | HSN1366   | HSN1366             |
| C | 131 | 19.4 | 64.7 | 163136 | 2  | AP005913  | AP005913 Oryza sat | C 204 | 19.2 | 63.3 | 32217  | 9  | AC017265  | AC017265 Drosophil  |
| C | 132 | 19.4 | 64.7 | 175727 | 10 | AC127580  | AC127580 Mus muscu | C 205 | 19.2 | 63.3 | 36993  | 2  | AC018007  | AC018007 Drosophil  |
| C | 133 | 19.4 | 64.7 | 179390 | 10 | AC122190  | AC122190 Mus muscu | C 206 | 19.2 | 63.3 | 39468  | 9  | AC007205  | AC007205 Homo sapi  |
| C | 134 | 19.4 | 64.7 | 181763 | 2  | AC136735  | AC136735 Mus muscu | C 207 | 19.2 | 63.3 | 42224  | 1  | MUCB373   | MUCB373             |
| C | 135 | 19.4 | 64.7 | 181917 | 2  | AC132728  | AC132728 Rattus no | C 208 | 19.2 | 63.3 | 67671  | 2  | AC102113  | AC102113 Mus muscu  |
| C | 136 | 19.4 | 64.7 | 183008 | 2  | AC019276  | AC019276 Homo sapi | C 209 | 19.2 | 63.3 | 69334  | 2  | AC101350  | AC101350 Mus muscu  |
| C | 137 | 19.4 | 64.7 | 185571 | 9  | AC133781  | AC133781 Homo sapi | C 210 | 19.2 | 63.3 | 106866 | 10 | AL732550  | AL732550 Mouse DNA  |
| C | 138 | 19.4 | 64.7 | 185921 | 2  | AC132645  | AC132645 Rattus no | C 211 | 19.2 | 63.3 | 110000 | 2  | AC096436  | AC096436 Rattus no  |

|     |      |      |        |    |              |                    |       |      |      |        |   |             |                    |
|-----|------|------|--------|----|--------------|--------------------|-------|------|------|--------|---|-------------|--------------------|
| 212 | 19   | 63.3 | 110000 | 2  | AC096436.1   | Continuation (2 of | 285   | 18.8 | 62.7 | 100426 | 9 | AL591063    | Human DNA          |
| 213 | 19   | 63.3 | 110000 | 2  | LMFLCHR36_23 | Continuation (24 o | c 286 | 18.8 | 62.7 | 102826 | 3 | AC136213    | Leishmani          |
| 214 | 19   | 63.3 | 129605 | 3  | AC006432     | AC006432 Drosophil | c 287 | 18.8 | 62.7 | 104770 | 9 | HS0245M18   | Human DNA          |
| 215 | 19   | 63.3 | 131135 | 2  | AC147551     | AC147551 Mus muscu | c 288 | 18.8 | 62.7 | 110000 | 2 | AC094316.1  | Continuation (2 of |
| 216 | 19   | 63.3 | 135713 | 2  | AC141030     | AC141030 Rattus no | c 289 | 18.8 | 62.7 | 110000 | 2 | LMFLCHR15_1 | Continuation (2 of |
| 217 | 19   | 63.3 | 136955 | 2  | AC011910     | AC011910 Drosophil | c 290 | 18.8 | 62.7 | 110000 | 2 | LMFLCHR26_0 | Continuation (15 o |
| 218 | 19   | 63.3 | 156667 | 2  | AC067911     | AC067911 Homo sapi | c 291 | 18.8 | 62.7 | 122850 | 9 | AL359744    | Human DNA          |
| 219 | 19   | 63.3 | 157637 | 9  | AC080106     | AC080106 Homo sapi | c 292 | 18.8 | 62.7 | 122850 | 9 | AL359744    | Human DNA          |
| 220 | 19   | 63.3 | 160597 | 9  | AL590666     | AL590666 Homo sapi | c 293 | 18.8 | 62.7 | 124456 | 2 | AC017830    | AC017830 Drosophil |
| 221 | 19   | 63.3 | 169404 | 9  | AL353751     | AL353751 Human DNA | c 294 | 18.8 | 62.7 | 138969 | 2 | AC060771    | Homo sapi          |
| 222 | 19   | 63.3 | 169420 | 2  | AC138641     | AC138641 Mus muscu | c 295 | 18.8 | 62.7 | 148230 | 2 | AC141548    | Rattus no          |
| 223 | 19   | 63.3 | 169420 | 3  | AC007416     | AC007416 Drosophil | c 296 | 18.8 | 62.7 | 148230 | 2 | AC141548    | Rattus no          |
| 224 | 19   | 63.3 | 173064 | 9  | AC022399     | AC022399 Homo sapi | c 297 | 18.8 | 62.7 | 151795 | 9 | HS85661     | Human DNA          |
| 225 | 19   | 63.3 | 173417 | 10 | AL607143     | AL607143 Mouse DNA | c 298 | 18.8 | 62.7 | 151795 | 9 | HS85661     | Human DNA          |
| 226 | 19   | 63.3 | 173735 | 2  | AC027063     | AC027063 Homo sapi | c 299 | 18.8 | 62.7 | 156360 | 2 | AC135360    | Mus muscu          |
| 227 | 19   | 63.3 | 176534 | 2  | AC135242     | AC135242 Oryza sat | c 300 | 18.8 | 62.7 | 158170 | 9 | AC016757    | Homo sapi          |
| 228 | 19   | 63.3 | 181445 | 9  | AC026367     | AC026367 Homo sapi |       |      |      |        |   | AC137072    | Genomic s          |
| 229 | 19   | 63.3 | 186123 | 2  | AC131987     | AC131987 Mus muscu |       |      |      |        |   |             |                    |
| 230 | 19   | 63.3 | 186563 | 9  | AC104452     | AC104452 Homo sapi |       |      |      |        |   |             |                    |
| 231 | 19   | 63.3 | 186925 | 2  | AC087503     | AC087503 Homo sapi |       |      |      |        |   |             |                    |
| 232 | 19   | 63.3 | 189514 | 2  | AC140803     | AC140803 Rattus no |       |      |      |        |   |             |                    |
| 233 | 19   | 63.3 | 195269 | 6  | AX417035     | AX417035 Sequence  |       |      |      |        |   |             |                    |
| 234 | 19   | 63.3 | 198024 | 9  | AP000893     | AP000893 Homo sapi |       |      |      |        |   |             |                    |
| 235 | 19   | 63.3 | 207743 | 4  | AC121066     | AC121066 Oryctolag |       |      |      |        |   |             |                    |
| 236 | 19   | 63.3 | 222284 | 2  | AC024442     | AC024442 Homo sapi |       |      |      |        |   |             |                    |
| 237 | 19   | 63.3 | 225068 | 2  | AC094495     | AC094495 Rattus no |       |      |      |        |   |             |                    |
| 238 | 19   | 63.3 | 226783 | 2  | AC108318     | AC108318 Rattus no |       |      |      |        |   |             |                    |
| 239 | 19   | 63.3 | 229315 | 2  | AC111736     | AC111736 Rattus no |       |      |      |        |   |             |                    |
| 240 | 19   | 63.3 | 234810 | 2  | AC105654     | AC105654 Rattus no |       |      |      |        |   |             |                    |
| 241 | 19   | 63.3 | 241509 | 2  | AC106608     | AC106608 Rattus no |       |      |      |        |   |             |                    |
| 242 | 19   | 63.3 | 259159 | 2  | AC111333     | AC111333 Rattus no |       |      |      |        |   |             |                    |
| 243 | 19   | 63.3 | 259314 | 2  | AC096024     | AC096024 Rattus no |       |      |      |        |   |             |                    |
| 244 | 19   | 63.3 | 260880 | 2  | AC134486     | AC134486 Rattus no |       |      |      |        |   |             |                    |
| 245 | 19   | 63.3 | 289516 | 3  | AE003619     | AE003619 Drosophil |       |      |      |        |   |             |                    |
| 246 | 19   | 63.3 | 291923 | 3  | AE003818     | AE003818 Drosophil |       |      |      |        |   |             |                    |
| 247 | 19   | 63.3 | 299130 | 1  | AP005077     | AP005077 Vibrio pa |       |      |      |        |   |             |                    |
| 248 | 19   | 63.3 | 333050 | 1  | AL596168     | AL596168 Listeria  |       |      |      |        |   |             |                    |
| 249 | 19   | 63.3 | 338450 | 1  | MLEPRTN4     | MLEPRTN4 Mycobacte |       |      |      |        |   |             |                    |
| 250 | 19   | 63.3 | 349980 | 6  | AX417044     | AX417044 Sequence  |       |      |      |        |   |             |                    |
| 251 | 18.8 | 62.7 | 689    | 3  | AF171868     | AF171868 Drosophil |       |      |      |        |   |             |                    |
| 252 | 18.8 | 62.7 | 1015   | 6  | BD275229     | BD275229 46 Human  |       |      |      |        |   |             |                    |
| 253 | 18.8 | 62.7 | 1104   | 8  | AF053553     | AF053553 Mesembrya |       |      |      |        |   |             |                    |
| 254 | 18.8 | 62.7 | 1143   | 6  | BD082130     | BD082130 Non-immun |       |      |      |        |   |             |                    |
| 255 | 18.8 | 62.7 | 1278   | 10 | BC005689     | BC005689 Mus muscu |       |      |      |        |   |             |                    |
| 256 | 18.8 | 62.7 | 1406   | 6  | BD275216     | BD275216 46 Human  |       |      |      |        |   |             |                    |
| 257 | 18.8 | 62.7 | 1500   | 6  | AX536902     | AX536902 Sequence  |       |      |      |        |   |             |                    |
| 258 | 18.8 | 62.7 | 1740   | 8  | AK061955     | AK061955 Oryza sat |       |      |      |        |   |             |                    |
| 259 | 18.8 | 62.7 | 1824   | 6  | I65510       | I65510 Sequence 1  |       |      |      |        |   |             |                    |
| 260 | 18.8 | 62.7 | 1874   | 5  | CHKHMGCOAS   | M60657 Chicken HMG |       |      |      |        |   |             |                    |
| 261 | 18.8 | 62.7 | 1969   | 8  | AK066380     | AK066380 Oryza sat |       |      |      |        |   |             |                    |
| 262 | 18.8 | 62.7 | 2198   | 10 | BC017154     | BC017154 Mus muscu |       |      |      |        |   |             |                    |
| 263 | 18.8 | 62.7 | 2212   | 10 | AB041537     | AB041537 Mus muscu |       |      |      |        |   |             |                    |
| 264 | 18.8 | 62.7 | 2270   | 10 | AK128920     | AK128920 Mus muscu |       |      |      |        |   |             |                    |
| 265 | 18.8 | 62.7 | 2907   | 6  | AK380508     | AK380508 Sequence  |       |      |      |        |   |             |                    |
| 266 | 18.8 | 62.7 | 2907   | 6  | AX332836     | AX332836 Sequence  |       |      |      |        |   |             |                    |
| 267 | 18.8 | 62.7 | 2907   | 6  | AX333341     | AX333341 Sequence  |       |      |      |        |   |             |                    |
| 268 | 18.8 | 62.7 | 2907   | 6  | AX333744     | AX333744 Sequence  |       |      |      |        |   |             |                    |
| 269 | 18.8 | 62.7 | 2907   | 6  | AX409671     | AX409671 Sequence  |       |      |      |        |   |             |                    |
| 270 | 18.8 | 62.7 | 2907   | 6  | HMCIP2EA     | M29873 Human cytoc |       |      |      |        |   |             |                    |
| 271 | 18.8 | 62.7 | 3000   | 9  | BC041174     | BC041174 Homo sapi |       |      |      |        |   |             |                    |
| 272 | 18.8 | 62.7 | 3119   | 9  | BC031945     | BC031945 Homo sapi |       |      |      |        |   |             |                    |
| 273 | 18.8 | 62.7 | 3708   | 8  | AY144944     | AY144944 Saccharom |       |      |      |        |   |             |                    |
| 274 | 18.8 | 62.7 | 4579   | 8  | GPU306607    | AJ306607 Gibberell |       |      |      |        |   |             |                    |
| 275 | 18.8 | 62.7 | 11189  | 1  | AE014091     | AE014091 Buchnera  |       |      |      |        |   |             |                    |
| 276 | 18.8 | 62.7 | 12580  | 8  | SPAC1296     | AL035439 S.pombe c |       |      |      |        |   |             |                    |
| 277 | 18.8 | 62.7 | 25065  | 8  | AC074098     | AC074098 Homo sapi |       |      |      |        |   |             |                    |
| 278 | 18.8 | 62.7 | 36522  | 9  | EX005019     | EX005019 Human DNA |       |      |      |        |   |             |                    |
| 279 | 18.8 | 62.7 | 54166  | 9  | AC126119     | AC126119 Homo sapi |       |      |      |        |   |             |                    |
| 280 | 18.8 | 62.7 | 66395  | 9  | HS119E23     | Z99570 Human DNA s |       |      |      |        |   |             |                    |
| 281 | 18.8 | 62.7 | 68755  | 9  | AC096559     | AC096559 Homo sapi |       |      |      |        |   |             |                    |
| 282 | 18.8 | 62.7 | 74685  | 2  | AC114682     | AC114682 Homo sapi |       |      |      |        |   |             |                    |
| 283 | 18.8 | 62.7 | 91679  | 2  | AC107984     | AC107984 Homo sapi |       |      |      |        |   |             |                    |
| 284 | 18.8 | 62.7 | 94516  | 8  | AF004526     | AF004526 Lotus cor |       |      |      |        |   |             |                    |

RESULT 1

AE006916/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
gene  
CDS

15619 bp DNA linear BCT 27-APR-2001  
Mycobacterium tuberculosis CDC1551, section 2 of 280 of the complete genome.  
AE006916 AE0000516  
AE006916.1 GI:13879055  
Mycobacterium tuberculosis CDC1551  
Mycobacterium tuberculosis CDC1551  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.  
1 (bases 1 to 15619)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains  
Unpublished  
2 (bases 1 to 15619)  
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.  
Direct Submission  
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers  
1. .15619  
/organism="Mycobacterium tuberculosis CDC1551"  
/mol\_type="genomic DNA"  
/strain="CDC1551"  
/db\_xref="taxon:83331"  
/note="clinical strain"  
78. .866  
/gene="MT0015"  
78. .866  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK44237.1"  
/db\_xref="GI:13879056"  
/translation="MELTPTCPENGETMIDRRSAMFSPVLVCLAGLLAATHG  
VSGGTIRSDAPVHGLVVRQAQSVNRLATREALTTRIDSVHGESVDTALAAMQR  
SKLAGVAANPVHGLVVTVDQAQRDANGFRPDASPDVLVHQDDTEAVNLAWN  
AGREAQMQDQRIIANSIARCVGNTLLNGRTYSPPTTAAIGDAAMQAALAAAPLV

```

gene
CDS
    TLYKQYVVRFGYGRYEEVHPDLOIVGYADPVRMHPAQAGPLDY"
    903..1601
    /gene="WT0016"
    903..1601
    /gene="WT0016"
    /note="similar to PID:1262359; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="glutamine amidotransferase, class I"
    /protein_id="AAK44238.1"
    /db_xref="GI:13879057"
    /translation="NRILVDNYSFVNLVQYLQGLGAEVNRNDHRLSDAAVA
    GQFDGVLSPGPGTPERAGASVIVHACAAHETPLGVCLGHQAIGVAFGATVDRAPE
    LLHGKTSVFTNNGVQLGDDPPTATRYHSLTILPKSLPAVLRVARTSGVIMAYO
    HTGLPIGVQFHPSILTEGGHRIANLWLTCCGWTQDQTLVRULENEVLTAISPHFT
    STASAGEATGRTSA"
    complement (1579..3459)
    /gene="WT0017"
    complement (1579..3459)
    /gene="WT0017"
    /note="similar to PID:2344821; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="serine/threonine protein kinase"
    /protein_id="AAK44239.1"
    /db_xref="GI:13879058"
    /translation="MTTSHLSDRYEIGLPGCMSEVHLARLRLHRDVAVKVLRA
    DLADPSYLRPREAQAALNHPAIVAVYDTGEAETPAGPLPIVVEYVDGVTLRD
    IVHTEGPTPKRAIEVIADACALNFHQNGIHRDKVPANIMISATNAVXMDFGIA
    RAIDAGNSVTQTAIVTAQYLPQEARQSDVDARSVSLGCVLYELTGEPPTG
    DSVSAVQHVREDPIPPSARHEGLSDLDVAVLKAKQNPENRYQTAAMRADLVRV
    HNGPEPAKVLTAERTSLLSAGNLSGPRTPDLPRLDDTDDRSISGVGRWA
    VVAVLAVTVVTVAINTEPGITRDVQVDPVRGSSADAIAIQLNRGFKIRTQKPS
    TIPPDHVLTQANTSVAGDEITVNVSTGPEOREIPDVSTLTIAEAVKLTAGFG
    RFKQNSFTPELVKVIQTNPPANQTSAINTVIIIVSGPATKQIPDVAGQTVDVA
    QKMLNVGTFKFSQASVDSPPRAGEVTGNPNAGTTVPVDSVIELQVSKGNQFVMDL
    SQMWVDAPRLALGWTGMDKGVADVAGGQENRVVYQNPAGTGVNRDGIITLRF
    GQ"
    complement (3456..4751)
    /gene="WT0018"
    complement (3456..4751)
    /gene="WT0018"
    /note="similar to PID:2344822; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="serine/threonine protein kinase"
    /protein_id="AAK44240.1"
    /db_xref="GI:13879059"
    /translation="MSPRVGTLSGRYLRQLRIATGGMQGWAEVDNRLGRRVAVKVL
    KSEFSDPEFTFRFAERTTAMLNHPGIAVSHVHDSQWNGEGRTAYLWELVNGEP
    LNSVLKTRGRLHALDMLQTRALQIAAGLVHRDVKPKNILITLVQVKITDF
    GIKAVDAAPTQGVMTQAYIAPEQALGHDSADSVYSLGVGVYEAWSGRKPPA
    GDGLATVAMKHIKSPPPPLDLPNNVRELIETLVNPNMRYRSGSPFADAVAVRA
    GRRPPRSTPPGERAPALIPSGTITRVAANSAGRTAASRRSPATGGRHPRRTFS
    SQKALUWAAGVIGALIAIVLVKAPGNSPQAPPTPTVTITGNPPASNTGGTDA
    SPRLNWERTGTRSHGLQSWVPTPHSRASLARYETAQ"
    complement (4748..6223)
    /gene="WT0019"
    complement (4748..6223)
    /gene="WT0019"
    /note="similar to GP:5139605; identified by sequence
    similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="penicillin-binding protein, putative"
    /protein_id="AAK44241.1"
    /db_xref="GI:13879060"
    /translation="MNASLRISVYTMALIVLLLNATNQVFTADGLADPNQRVL
    LDEYSRQQTACGQLLAVSDTGRFLRVFPNPNVAPVTFGTSLSRYSIALER
    AEDPILNGSDRLFRGLADFTGRDPRGNVDTTINPRTQQAGWAMQOQGCYPCKG

```

```

AAVALLPSTGKILALVSSSDPNLLASHNPEVQAQWRLGDNFASPLTNRAISETY
PGSTFKVITTAALAAAGATEQUATAPTIPLPGSTAGLENYGAPCDEFTVSLRE
AFVKSNTAFVQIGIRTGADARLMARAPFLDPPRPTPLQVAESTVGPPIPSAALGM
TSIGCQVALTFLANAEIATANGITMRPLDVLGSLKGPDLANI-STTVGQORRAVS
POVAKILTELMVGAQKQKGAIPGVQIASKTGTAEHTDPRTPPHAWYIAFAPAQ
APKAVAVLNVGADRLSATGALAAPIGRAVIEAALQGEF"
complement (6220..7629)
/gene="WT0020"
complement (6220..7629)
/gene="WT0020"
/note="similar to GB:D10483 SP:P16457 GB:M30807 PID:146039
PID:216503; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cell division protein FtsW"
/protein_id="AAK44242.1"
/db_xref="GI:13879061"
/translation="MTRLQAPVAVTPELPTRRNAELLLCFAAVITFAALLVQANQ
DOGVMDLTSYGLAFLTLFGSAHLAIRFPAPYTDPLLLPVVALLNGLGLVIMHRLDLV
DNEIGHRHPSANQOMLTLGVAAAFALVWTFKDRHOLARVGYICGLAGLVFLAVPA
LLPALSQNGAKINIRLPGFSIQPAESKILL-LFFSAVLVAKEGLTSAGKHLGCM
TLPFRDLAPLAAWVISGVNVFEDKGLGASLLLTSLVVVYLATQRFWSVYIGLTL
FAAGTLVAYFIEHVRVLRVQTLWDFADPDGTGYQIVQSLFSFATGIFGTLGNGQP
DTVPAASDTFIIAFAFEEGLVGLTALMLTYTIVIRGLRTAIRTDRSFGKLLAAGLS
STLAIQLFTIVGVGTRLLPLTGLTTPWMSYSGSSLLANYILLAILARISHGARREPT
RPNKSPITAACTEVIERY"
complement (7626..9161)
/gene="WT0021"
complement (7626..9161)
/gene="WT0021"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44243.1"
/db_xref="GI:13879062"
/translation="MTLVRYAASDRGLVRANNEDSVYAGARLLALADMGCHHAAGE
VASQIVLAHLHLDDEPGDOLLAKLDAARAGNSAIAAQQEMEPDLEGMTTITAIL
FAGNLRLVHIGDSRGYLLRDELQITKDDTFTVQTLVDEGRITPEEAHSPQSLIM
KALGHEVPTLTMEARAGRYLLCSDPLSDPSDETILBALQIPEVAESAHRLIEL
ALRGGQPNVTWVADVVDYDQOTQPILAGVSGDDDLTLPTNATGSAISORKE
IVKRVPPQADTFESRPSGRBLAFVVALVTVLMTAGLIGRAIIRSNVYVADYAGSVS
INRGIGSLLMSLQPYLMGCLSPRNLSQISQCSGGLDCHLMKLEDRPPEAQA
VRAPAGLTDDAICQLRELAANSLPCCPAPRATSPGPRAPPTTSETTBNVTSSP
ASPSTTSASAPTGTTPATPSASPAASPPTPMTPTTSPTMAALPPPPPPGIDCR
AAA"
complement (9259..9726)
/gene="WT0022"
complement (9259..9726)
/gene="WT0022"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44244.1"
/db_xref="GI:13879063"
/translation="MQGLVQLTRAGFLMLLVFWFWSVLRLKTDIYAPTGAUMVRRG
LALRGLGAGQRRAHRLVVTGALTGARITLSEQVLLGRADDSLTVDITDFAST
RHARLSMGSEWYEDLGSTNGTYLDRAKVTTAVRVPTGTPTVTRIGKTAIELRP"
complement (9850..11415)
/gene="WT0023"
complement (9850..11415)
/gene="WT0023"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44245.1"
/db_xref="GI:13879064"
/translation="MGSOKRLVORVERKLEOTVGDAFARIFGGSIVPOVEALLRREA
ADGTSLOGNRLAPANEVITILGVHDFEKLADPELAKSTGARNLADYIQOQWQYVC
DVVRFEGSSNLHTQFARGTVPNDVETHEPFDICARFQSNHAFGAGFAPVMSDNS
SYRGQGGQGRFDEYDRIYARYPQEDPRGGPDQGGSDPRGGYPPETGGYPPQGYPRP

```





Hinxton, Cambridge CB10 1SA, UK. P74 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France. Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France.

Location/Qualifiers

1. 343050

/organism="Mycobacterium bovis subsp. bovis AF2122/97"

/mol\_type="genomic DNA"

/strain="AF2122/97"

/db\_xref="taxon:233413"

1. 1524

/locus\_tag="dnaA"

/locus\_tag="Mb0001"

1. 1524

/gene="dnaA"

/locus\_tag="Mb0001"

/note="Mb0001", dnaA, len: 507 aa. Equivalent to Rv0001, len: 507 aa, from Mycobacterium tuberculosis strain H37Rv, (99.6% identity in 507 aa overlap). dnaA, chromosomal replication initiator protein (see citations below), equivalent to other Mycobacterial CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g. P46388|DNAA\_MYCE from Mycobacterium leprae (502 aa); Q9L7L7|DNAA\_MYCEA from Mycobacterium paratuberculosis (509 aa); P49990|DNAA\_MYCAV from Mycobacterium avium (508 aa); P49992|DNAA\_MYCSM from Mycobacterium smegmatis (504 aa); etc. Also highly similar to others except in N-terminus e.g. Q9ZH75|DNAA\_STRCH CHROMOSOMAL REPLICATION INITIATOR PROTEIN from Streptomyces chrysosallus (624 aa); Q9ZH76|DNAA\_STRRE from Streptomyces reticuli (643 aa); DNAA\_ECOLI P03004|B3702 chromosomal replication initiator protein from Escherichia coli strain K12 (467 aa), FASTA scores: opt: 966, E(): 0, 43.2% identity in 389 aa overlap; etc. Contains PS00017 ATP/GTP-binding site motif A (p-loop) and PS01008 DnaA protein signature. BELONGS TO THE DNAA FAMILY. Note that the first base of this gene has been taken as base 1 of the Mycobacterium bovis genomic sequence."

/codon\_start=1

/evidence=experimental

/transl\_table=11

/product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA"

/protein\_id="CAD92863.1"

/db\_xref="GI:31616763"

/translation="MTDDPGSGFTTVNAVVSSELNGPKVDDGSPSSDANLSAPLTQQ RANLVQPTIVGFAALLVPFSFQNEIHLRAFTIDALRRRLGHQQLGVRAP PATDEADTTVPSENATTPSDTTNDNDIDPSAAAGNDQNSWPSYFPRNTDS ATAGVTSLNRRYTFDTVIGASNFHAALAAEAPARNPLFTWGESGLGKTHLL HAAGNVAQRLPFGVRVVKVYNSDPPKOLATLEDRLTRPFWGLITDVQPI EGKGIQEEPPHTLHNANKQIVISDPPKOLATLEDRLTRPFWGLITDVQPI LETRIALRKAQMERLAIEDDULELIASSIERNIRLEGALIRVTAFASLNTPDK ALAEIVRLDIADANTWQISAAITMAATAEFDTTVELSGPKTRALAQSRQIAYL CRELTDLSPKIGQAFGRDHTTVMYAQKILSENAERREVFHVKEITLIRQSRK" 2052..3260

/gene="dnaA"

/locus\_tag="Mb0002"

2052..3260

/locus\_tag="dnaA"

/gene="dnaA"

/EC\_number="2.7.7.7"

/note="Mb0002", dnaA, len: 402 aa. Equivalent to Rv0002, len: 402 aa, from Mycobacterium tuberculosis strain H37Rv, (99.8% identity in 402 aa overlap). dnaA, DNA polymerase III (beta chain) (EC 2.7.7.7) (see citations below), equivalent to other Mycobacterial DNA POLYMERASES III BETA CHAIN e.g. NP\_301130.1|NC\_002677 from Mycobacterium leprae (399 aa); Q9L7L6|DP3B\_MYCEA from Mycobacterium avium subsp. paratuberculosis (399 aa); P52851|DP3B\_MYCSM from Mycobacterium smegmatis (397 aa); etc. Also highly similar to others e.g. P27903|DP3B\_STRCO DNA POLYMERASE III BETA CHAIN from Streptomyces coelicolor (376 aa), FASTA scores: opt: 1189, E(): 0, (52.8% identity in 337 aa overlap); P21174|DP3B\_MICLU from Micrococcus luteus (310 aa); P52023|DP3B\_SYN7 from Synecococcus sp. strain PCC 7942

(375 aa); etc. Overlaps and extends CDS in neighbouring cosmid MTCY10H4.01."

/codon\_start=1

/evidence=experimental

/transl\_table=11

/product="DNA POLYMERASE III (BETA CHAIN) DNAN (DNA NUCLEOTIDYLTRANSFERASE)"

/protein\_id="CAD92864.1"

/db\_xref="GI:31616764"

/translation="MDAATTVGLTDLTFRLLRESFADAVSWAKNLPARPAVPLSG VLLTGSUNGLTISGFDYVSAEQVAGIEVSPGVLVSGRLSLDITRALPNKPVGVHV EGNVALTCGNARSLPTMPVEDYPTLTLEETGLLPAELFAEISOVAIAAGRDDIT LFMLTGIRVEILGETVVLAAADRFLAVRELKWSASSPDIEAAVLVPKATLAEAKAG IGGSDVRLSLGTGPGVKGGLGSGNGKRSRTLLRDLAEFPKFKQLLTPHTVAIMD VAELEIAIKLVADVADGAGVQVMEFAGSVLSAGADDVVAEEDLVVDVYAGEPLTIA FNPTVLTDLGLSLSRSERVSFGFTTAGKPAALLRPVSGDDRPVAGLNGNGPFFAVSTDYV YLLMPVRLPG" 3280..4437

/gene="recF"

/locus\_tag="Mb0003"

3280..4437

/gene="recF"

/locus\_tag="Mb0003"

/note="Mb0003", recF, len: 385 aa. Equivalent to Rv0003, len: 385 aa, from Mycobacterium tuberculosis strain H37Rv, (99.5% identity in 385 aa overlap). recF, DNA replication and repair protein (see citations below), equivalent to other Mycobacterial DNA replication and repair proteins e.g. NP\_301131.1|NC\_002677 from Mycobacterium leprae (385 aa); Q9L7L5|REC\_F\_MYCEA from Mycobacterium avium subsp. paratuberculosis (385 aa); P50916|REC\_F\_MYCSM from Mycobacterium smegmatis (384 aa); etc. Also highly similar to others e.g. P36176|REC\_F\_STRCO DNA REPLICATION AND REPAIR PROTEIN from Streptomyces coelicolor (373 aa); REPAIR PROTEIN from Synecocystis sp. strain PCC 6803 (384 aa); NP\_469352.1|NC\_003212 from Listeria innocua (370 aa); etc. Contains PS00017 ATP/GTP-binding site motif A (p-loop), PS00617 recF protein signature 1, and PS00618 recF protein signature 2. BELONGS TO THE REC\_F FAMILY."

/codon\_start=1

/evidence=experimental

/transl\_table=11

/product="DNA REPLICATION AND REPAIR PROTEIN REC\_F (SINGLE-STRAND DNA BINDING PROTEIN)"

/protein\_id="CAD92865.1"

/db\_xref="GI:31616765"

/translation="MIVRHIGURDRFRWACVDLELHPGRTVTVGNGYKTNLLIALM YETTLGSHRVSADLEPLIRVGTDRIVSTIVNDGECVDEIAIGRVNKAELNRSV RSTRDVGVLRVLAFLPAEDGLVIRGDPADRRYLDLAIIVREPAVAARVRYVRQ RTALLKSVPCARYGRGVDFDTLEVWDSRLAEHGAELVAARLDLVNQLAPEYKAYQL LAPERSASIGYRASMDVTGPSEQDTRQILAAALLAALAAARDAELRGVCLGVGPH RDLLRLRGDQPAKGAFSGEASLAVRLAAYOLLRVDDGPEPVLDDVFAELDM RRLAATAESAEQVLTVAVLDPAGHDARVIDRADDTGSMVSLP" 4434..4997

/locus\_tag="Mb0004"

4434..4997

/locus\_tag="Mb0004"

/note="Mb0004", len: 187 aa. Equivalent to Rv0004, len: 187 aa, from Mycobacterium tuberculosis strain H37Rv, (99.5% identity in 187 aa overlap). Conserved hypothetical protein (see citation below), highly similar, but longer 21 aa in N-terminus, to AAF33696.1|AF222789 unknown protein from Mycobacterium avium subsp. paratuberculosis (166 aa); and highly similar to NP\_301132.1|NC\_002677 conserved hypothetical protein from Mycobacterium leprae (189 aa); S70990 hypothetical protein from Mycobacterium smegmatis (194 aa). Also similar to in C-terminus to C-terminal part of P35925|YREG\_STRCO HYPOTHETICAL 19.8 KDA PROTEIN (IN REC\_F-GENE INTERGENIC REGION) from Streptomyces coelicolor (190 aa), FASTA scores: opt: 404, E(): 3.9e-18, (40.7% identity in 189 aa overlap)."

/codon\_start=1

/transl\_table=11



```

/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD92866.1"
/db_xref="GI:31616766"
/translation="MTGSDVRPDQNGRGLMKSPGLDVRITLDEARAAARQCQDAG
RGVAVASGRVAGRRWSGPGDIRDQPLGKAARELAKRGWSVRVAGMVLGOW
SAVVGQIAEHARPTALNDGVLSTVAESTAWATQLRINQAOLAKIAAAVNDVRSI
KITGPAAPSWRKGRPHIAGRGPRDTYG"
gene          5123..7267
/gene="gyrB"
/locus_tag="Mb00005"
CDS           5123..7267
/gene="gyrB"
/locus_tag="Mb00005"
/EC_number="5.99.1.3"
/notes="Mb00005. gyrB, len: 714 aa. Equivalent to Rv0005,
len: 714 aa, from Mycobacterium tuberculosis strain H37RV,
(99.9% identity in 714 aa overlap). gyrB, DNA gyrase
subunit B (EC 5.99.1.3) (see citations below), equivalent,
except in N-terminus, to other Mycobacterial DNA GYRASES
SUBUNIT B e.g. T10005 from Mycobacterium leprae (697 aa);
Q9L7L3|GYRB MYCPA from Mycobacterium avium subsp.
paratuberculosis (677 aa) (has its N-terminus shorter);
F48355|GYRB MYCSM from Mycobacterium smegmatis (675 aa);
etc. Also highly similar to others e.g. T10969 from
Streptomyces coelicolor (686 aa); P50075|GYBS_STRSH from
Query Match      100.0%; Score 30; DB 1; Length 343050;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGAGAACGAGGATGACCAATGAG 30
Db 24985 TGGCGAGAACGAGGATGACCAATGAG 24956

RESULT 5
MCLB1770/c
LOCUS          MCLB1770          37821 bp          DNA          BCT 29-AUG-1997
DEFINITION    Mycobacterium leprae cosmid B1770.
ACCESSION     270722
VERSION       270722.1 GI:2344819
KEYWORDS      Anthranilate synthase; DNA A; DNA gyrase; DnaN; intein; PBP;
              peptidyl-prolyl isomerase; pseudogene; RecF; RodA; Ser, Thr-protein
              kinases; tRNA-Ala; tRNA-Ile; tRNA-Leu.
SOURCE        Mycobacterium leprae
ORGANISM      Mycobacterium leprae
              Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium.
REFERENCE     1 (bases 1 to 37821)
AUTHORS       Fsihi.H., De Rossi.E., Salazar.L., Cantoni.R., Labo.M.,
              Riccardi.G., Takiff.H.E., Eigmeier.K., Bergh.S. and Cole.S.T.
              Gene arrangement and organization in a approximately 76 kb fragment
              encompassing the oric region of the chromosome of Mycobacterium
              leprae
TITLE         Use of an ordered cosmid library to deduce the genomic organization
              of Mycobacterium leprae
JOURNAL       Mol. Microbiol. 7 (2), 197-206 (1993)
MEDLINE       93188700
PUBMED        8446027
REFERENCE     2 (bases 1 to 37821)
AUTHORS       Fsihi.H., De Rossi.E., Salazar.L., Cantoni.R., Labo.M.,
              Riccardi.G., Takiff.H.E., Eigmeier.K., Bergh.S. and Cole.S.T.
              Gene arrangement and organization in a approximately 76 kb fragment
              encompassing the oric region of the chromosome of Mycobacterium
              leprae
TITLE         Homing events in the gyrA gene of some mycobacteria
JOURNAL       Proc. Natl. Acad. Sci. U.S.A. 93 (8), 3410-3415 (1996)
MEDLINE       96194983
PUBMED        8622949
REFERENCE     4 (bases 1 to 37821)
AUTHORS       Salazar.L., Fsihi.H., de Rossi.E., Riccardi.G., Rios.C., Cole.S.T.
              and Takiff.H.E.
              Organization of the origins of replication of the chromosomes of
              Mycobacterium smegmatis, Mycobacterium leprae and Mycobacterium

```

```

tuberculosis and isolation of a functional origin from M. smegmatis
Mol. Microbiol. 20 (2), 283-293 (1996)
96310367
8733228
5 (bases 1 to 37821)
Cole.S.T.
Direct Submission
Submitted (25-AUG-1997) Dr. Stewart T. Cole, Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France Requests for cosmids should be sent to
Karin Eigmeier (kei@pasteur.fr)
On Aug 27, 1997 this sequence version replaced gi:1262351.
Notes:
Work in Paris is supported by the Heiser Trust, the Association
Francaise Raoul Follereau and the Groupement de Recherches et des
Etudes des Genomes (GIP-GREG).
CDS are numbered using the following system eg MLCB33.01c. ML (M.
leprae), CB33 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. All CDS
over 100 codons have been analysed. Gene prediction is based on
positional base preference in codons especially where there is an
increase in the observed/expected third position G + C. CAUTION:
We may not have predicted the correct initiation codon. Where
possible we choose an initiation codon (atg, gtg, or ttg) which is
preceded by an upstream ribosome binding site sequence (optimally
5-13bp before the initiation codon). If this cannot be identified
we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid B1770
overlaps cosmids L222 and B628 at the 5' and 3'-ends, respectively.
Location/Qualifiers
1..37821
/organism="Mycobacterium leprae"
/mol_type="genomic DNA"
/db_xref="taxon:1769"
/clone="cosmid B1770"
1..598
/gene="dnaA"
<1..598
/gene="dnaA"
/notes="MLB1770.01; dnaA"; The 5' part of the M. leprae
dnaA gene is on cosmid L222 (accession number L39923)"
/codon_start=2
/transl_table=11
/product="DnaA"
/protein_id="CAA94708.1"
/db_xref="GI:1262352"
/translation="QLATLEDRRLTRFENGILTDVQPELETRAILRKKAKMERLAV
PDVUELLASSIERNELEGALIRVTAFSLNKTAIDKALAEIVRLDLIADASTMQI
SAATIMTATFYDTIEELRPGKTRALAQSRQIAMYLCRELTDLSLPKTGQAFGRD
HTVMYAKRIKLSMAERREVDHVHVELTTRIRQSKK"
1113..2312
/gene="dnaN"
1113..2312
/gene="dnaN"
/notes="MLB1770.02; dnaN; len: 399"
/codon_start=1
/transl_table=11
/product="DnaN"
/protein_id="CAA94709.1"
/db_xref="GI:1262353"
/db_xref="GOA:P46387"
/db_xref="SWISS-PROT:P46387"
/translation="MDLAKTNVGCSDLKFLCLARESFASVSWAKYLPTRPTVPLSG

```



VLTGSDSGLTGSDYEVSAEYVQVAEIASGSSVLVSGRLSDITRALPNKPVHYV  
DGNRAVLCGSARFSLTMAVEDYPTLPTDPTGTPSDFAEAGVQVAIAAGRDYT  
LPMLTGIRIEIGDVTVAATDFRLAVRELKWSVLSDFEASVLVPAKTLVEVAKAG  
TDGSGVCLSGAGVGKDLGFI SGGKRSYTRLLDADFPPKROLLPAETHVAVATID  
FENYLTIDGLASVHSERVSFGFTTSPKALLRPTSDNDVHPHTDGFPPALPDYVILL  
MPVRLFG"  
2313..3470  
/gene="recF"  
2313..3470  
/gene="recF"  
/note="MLB1770.03, recF; len: 385"  
/codon\_start=1  
/transl\_table=11  
/product="RecF"  
/protein\_id="CAA94710.1"  
/db\_xref="GI:1262354"  
/db\_xref="GOA:P46391"  
/db\_xref="SWISS-PROT:P46391"  
/translations="NTVRRHGLRDRPSWDHVDLELNPGRVTFPGNGKXNLIEALV  
YSTLSHRVGTIDIPTRAGTIRAVSTIVNRECAIDLEIAGRANRANRLNSLV  
RGNREVGLRVAVLFAPELDALVCGDPANRRYLDLATVQPVTAARVADYKVLQ  
RLALKSLAARYSRDQGLDLDVWDTRIAEHGAELMAARIDLNVQLAPEYKAYOL  
LAPGRTASTSYRASLDIGGIAGVSGSDRALLOADLLAGLSTRNVELERGI CLVGP  
RDELRGLGPQAKGPASGHGSLSAIALRLAAYELLRADGNPEVLLDDVFAELDA  
RCALATVAESAQVLVTSNAQEDIPVGNDAKWTVVDLRSDSGRVSVPY"  
3454..3459  
/gene="recF"  
/notes="possible RBS, GGTGG, for MLB1770.04"  
3467..4066  
/gene="MLB1770.04"  
3467..4066  
/gene="MLB1770.04"  
/note="MLB1770.04, unknown, previously orf199; len: 199 ;  
encodes hypothetical protein in the RecF-GyrB intergenic  
region"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="CAA94711.1"  
/db\_xref="GI:1262355"  
/db\_xref="SWISS-PROT:Q8CDF4"  
/translations="MIESNESYGGDTIEPLTGLSGFDLVRALBEARAAACAQGD  
AGSHVPPVPPFVTRRNWSPGPDVDPQLGVANDLAKRQWSQVAEGRVFG  
QMASVGGQADAHAFVGLNGVLSYTAESTAWATOLRIMOQALLAKIAAAVGVVVT  
SLKLTGPTASRKGFWHLIAGRGPRTTASRHAGSELSR"  
4195..4198  
/note="possible RBS, GGAG, for gyrB"  
4203..6296  
/gene="gyrB"  
4203..6296  
/gene="gyrB"  
/note="MLB1770.05, gyrB; len: = 697"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase B"  
/protein\_id="CAA94712.1"  
/db\_xref="GI:1262356"  
/db\_xref="GOA:Q59533"  
/db\_xref="SWISS-PROT:Q59533"  
/translations="MTAAVTGPTCLNKLKESIQVAAQKRAQDEYGAASITILEGLAV  
RKPGMYVGSTGERGHLHLEWVDNVDMAGYATQVDVRLFDGSDVADNGRGI  
PVAHVATGVTQVMTQKLAGGKFGSDGVSGLGVGVSVVNLSTREVDIK  
RDGYMSQFYDKAVPGILKQGEATEATGTRIFWADPDIPETTKYDFGVARRIQEVA  
PLNGKLTINLVDKQDEVDVDSVTAEPVAMTVEEKSTESSAPFKVRHRTFHP  
GGLAVDFVHNTRKTPQQSIIDFGKGAGHEVEVAMQWNGYSESVHVFANTINHE  
GKTHGEGFRALSVNRYAKDKLLKQDPNLTDGDIRGLAALVSVKVSPEQFSGV  
TKTKLGNTEKVSQVRVNCQHLHFEANPDVAVVNAKIAISAQAKIARAKRGLQR  
RKSNADIGLPGKLADCRDTPRSSELYVVEGDSAGSAGSDMSFQALPLRGKII  
NVKARIDVLNKEVQAIITALGTGTHDFDISRLRYHKIVLMADADVQGHISTLL  
LTLFLRMRLIEHGVFLAQPPLYKLMQWMDPEFAYSDESDRDLGLETGLKLGKIN  
KEDGIQYKGLGMDAKELWETTMDFPSVRLRVQVTLDDAAAADLSEILMGEDVDAR  
SFTRNAKDVRFIDV"

6340..6345  
/note="possible RBS, GAGGAA, for gyrA"  
6349..10170  
/gene="gyrA"  
6349..10170  
/gene="gyrA"  
/note="MLB1770.06, gyrA, encodes a GyrA precursor protein  
spliced into a mature A subunit of DNA gyrase and a free  
protein intron (intein); len: 1273"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase A, and intein"  
/protein\_id="CAA94713.1"  
/db\_xref="GI:1262357"  
/db\_xref="GOA:Q57532"  
/db\_xref="SWISS-PROT:Q57532"  
/translations="MTDITLPPGDS:QRVSPVDIQQMORSY:DIYAMSVIVGRALPE  
VRDLKPKVHRRVLYAMLDGFRDRSHAKSARSVAETMGNYHPHGDSIYDTLVRMAQ  
Query Match 79.3%; Score 23.8; DB 1; Length 37821;  
Best Local Similarity 92.6%; Pred. No.12;  
Matches 25; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 GGCCAGAACCCAGGAGTAGCACCACATGA 29  
|||||  
Db 27225 GGCCAGAACCCAGGAGTAGCACCACATGA 27199  
|||||  
RESULT 6  
MLEPRTN1/c  
LOCUS MLEPRTN1 344050 bp DNA linear BCT 20-FEB-2001  
DEFINITION Mycobacterium leprae strain TN complete genome; segment 1/10.  
ACCESSION AL583917.1  
VERSION AL583917.1 GI:13092412  
KEYWORDS  
SOURCE Mycobacterium leprae  
ORGANISM Mycobacterium leprae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
REFERENCE 1 (bases 1 to 344050)  
AUTHORS Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,  
Wheeler, P.R., Honore, N., Ganier, T., Churcher, C., Harris, D.,  
Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,  
Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A.,  
Hamin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C.,  
Maclean, J., Moule, S., Murphy, L., Oliver, Quail, M.A.,  
Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S.,  
Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K.,  
Taylor, K., Whitehead, S., Woodward, J.R. and Barrell, B.G.  
TITLE Massive gene decay in the leprosy bacillus  
JOURNAL Nature 409 (6823), 1007-1011 (2001)  
MEDLINE 21128732  
PUBMED 11234002  
REFERENCE 2 (bases 1 to 344050)  
AUTHORS Parkhill, J.  
TITLE Direct Submission  
JOURNAL Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium  
leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome  
Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique  
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,  
75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk  
NOTES:  
Details of M. leprae sequencing at the Sanger Centre are available  
from [http://www.sanger.ac.uk/Projects/M\\_leprae/](http://www.sanger.ac.uk/Projects/M_leprae/) A relational  
database containing the M. leprae sequences is available from  
<http://genolist.pasteur.fr/Leprama/>.  
FEATURES  
source  
1..344050  
/organism="Mycobacterium leprae"  
/mol\_type="Genomic DNA"  
/strain="TN"  
/db\_xref="taxon:1769"  
1..1566  
gene

```

/gene="dnaA"
/notes="synonym: ML0001"
1..1566
/gene="dnaA"
/notes="Similar to M. tuberculosis dnaA, chromosomal
replication initiator protein, SW:DNAA_MYCTU (P49593) (507
aa); Fasta score E(): 0, 87.2% identity in 507 aa overlap
and Mycobacterium smegmatis dnaA, SW:DNAA_MYCSM (P49992)
(495 aa); Fasta score E(): 0, 69.2% identity in 441 aa
overlap. Previously sequenced as SW:DNAA_MYCLE (P46388)
(521 aa); Fasta score E(): 0, 99.8% identity in 521 aa
overlap. Contains Pfam match to entry PF00308 bac_dnaA,
Bacterial dnaA protein. Contains PS00017 ATP/GTP-binding
site motif A (P-loop). Contains PS01008 DnaA protein
signature."
/codon_start=1
/transl_table=11
/product="putative chromosomal replication initiator
protein"
/protein_id="CAC29509.1"
/db_xref="GI:13092413"
/db_xref="GOA:P46388"
/db_xref="SWISS-PROT:P46388"
/translation="MFVPHAKKPEIYENQDRTSLADLSLFGTTVMNAVSELNGESN
TDDATNDSTLVLTPOQRAWNLVQPLTIIIEGFALLSVPSVQVNEIERHLRTPIT
DALSRRLQOQLGVRIAPSTHDIDNSSADVLLTDDCGTDTDNYGEPRLTGEYCG
LPTVTERPHHTSTVGSPLARVTFEFVIGASNEFAHAALAEAPAPAYNPL
FIWSEGLKHLHAAAGNVAORLPFGMRVKYVNSTEFTNDFNSLNDKRVAPKRSY
RDVLLVDDIQTIEGKEGQEEFFHTNLNANKQIVISSDRPPKQATLEDRLT
RFENGLTDVQPPLETRIIARLKQMERLAVPGDVELELIASSIRNRELEGALIR
VTAFAKNTAIDKALAEIIVRLDIADASTWQISAAITMTATAEYFTTIEELRGFGK
TRALAQKQIYAMLYCRELTDLSLPKIQAGFRDHTTWYAKRILSEMAERREVFQHV
KELTRIRORSKR"
544..1482
/gene="dnaA"
/notes="Pfam match to entry PF00308 bac_dnaA, Bacterial
dnaA protein, score 712.20, E-value 8.3e-240"
664..687
/gene="dnaA"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
1426..1482
/gene="dnaA"
/notes="PS01008 DnaA protein signature"
2081..3280
/gene="dnaA"
/notes="synonym: ML0002"
2081..3280
/gene="dnaA"
/EC-number="2.7.7.7"
/notes="Similar to M. tuberculosis dnaA, DNA polymerase
III, [beta] subunit, SW:DP3B_MYCTU (Q50790) (402 aa);
Fasta score E(): 0, 80.9% identity in 403 aa overlap and
Mycobacterium smegmatis dnaA, SW:DP3B_MYCSM (P52851) (397
aa); Fasta score E(): 0, 77.3% identity in 397 aa overlap.
Previously sequenced as SW:DP3B_MYCLE (P46387) (399 aa);
Fasta score E(): 0, 100.0% identity in 399 aa overlap.
Contains Pfam match to entry PF00712 DNA_pol3_beta, DNA
polymerase III beta subunit."
/codon_start=1
/transl_table=11
/product="putative DNA polymerase III, [beta] subunit"
/protein_id="CAC29510.1"
/db_xref="GI:13092414"
/db_xref="GOA:P46387"
/db_xref="SWISS-PROT:P46387"
/translation="MDIAKTVNGSDLKFKCLARESFASVSWVKYLPTRPTVPVLSG
VLLTGSGLTISGDFVEAIVGVAEIASSGVLVSGRLSDITRALPKVHPYV
DGNVALTCCSARESLPTMAVEDYPTLPTLPDRTGTLPDVFEEAIGOVAIAGRDYV
LPMLTGTIRISGDTVLVAATDRPLAVRELKMSVLSSDPEASVLPVAKTJAEVAKG
TDGSGVCLISGKGVGKDGILFISGGGKSTRLRLDAEPFKRQLLPHEHTAVATID
VAENTEAIKVALVADRGAVQRMFEFGDILRLSAGADVGRAEEDLAVAPTEPLTIA
FNENPLYTDGLASVHSERSVFGFTTSPKALLRPTSDNDVHPTHDGFPALPTDYVLL
MPVRLPG"

```

```

2117..3214
/gene="dnaA"
/notes="Pfam match to entry PF00712 DNA_pol3_beta, DNA
polymerase III beta subunit, score 592.70, E-value
2.2e-174"
3281..4438
/gene="recP"
/notes="synonym: ML0003"
3281..4438
/gene="recP"
/notes="Similar to M. tuberculosis recP, DNA replication
and SOS induction protein, SW:RECF_MYCTU (Q59586) (385
aa); Fasta score E(): 0, 76.4% identity in 385 aa overlap
and Mycobacterium smegmatis recP, SW:RECF_MYCSM (P50916)
(384 aa); Fasta score E(): 0, 70.0% identity in 383 aa
overlap. Previously sequenced as SW:RECF_MYCLE (P46391)
(385 aa); Fasta score E(): 0, 99.5% identity in 385 aa
overlap. Contains 2 Pfam matches to entry PF00470 RecF,
RecF protein. Contains PS00617 RecF protein signature 1.
Contains PS00017 ATP/GTP-binding site motif A (P-loop).
Contains PS00618 RecF protein signature 2."
/codon_start=1
/transl_table=11
/product="putative DNA replication and SOS induction
protein"
/protein_id="CAC29511.1"
/db_xref="GI:13092415"
/db_xref="GOA:P46391"
/db_xref="SWISS-PROT:P46391"
/translation="MYVHFGCLRDFRSDHVDLELNPRTVFFGPNNGKNTLIEALW
XSTLSSHRVGTIDPLIRAGTIRAIIVTVNREGRECAIDLEIAAGRANRLNRSIV
KMTRVGVGLRAVLPAPEDLALVCGDPAARRVYLLDIAVTQPVIAAVRADYDKVLRQ
RTALKLSAAAYRSQDQGLDVLVDWDTRLAEHGAELMAARI DLVNOLABEVEAYOL
LAPCSRTASISYRASLDIGGIAGYSSSDRLADLQDLLAGLSTRENVEISGICVAGH
RDEELRUGDQPAKGFASHGNSIALRLAAYELLRADNENFVLLDDVFAELDAA
RCRLATVAESAQVLTSAQAQDIPVGWDARVTVLDLSDSDSGRVSVVP"
3287..3415
/gene="recP"
/notes="Pfam match to entry PF00470 RecF, RecF protein,
score 18.90, E-value 0.00024"
3368..3391
/gene="recP"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
3614..3691
/gene="recP"
/notes="PS00617 RecF protein signature 1"
4100..4423
/gene="recP"
/notes="Pfam match to entry PF00470 RecF, RecF protein,
score 5.20, E-value 1.5"
4244..4297
/gene="recP"
/notes="PS00618 RecF protein signature 2"
4435..5004
/gene="ML0004"
/notes="ML0004"
/notes="Similar to hypothetical proteins from mycobacteria
e.g. M. tuberculosis RV0004, hypothetical protein,
TR:PT1573 (EMBL:AL123456) (187 aa); Fasta score E(): 0,
77.2% identity in 167 aa overlap. Previously sequenced as
TR:050181 (EMBL:Z70722) (199 aa); Fasta score E(): 0,
99.5% identity in 187 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC29512.1"
/db_xref="GI:13092416"
/db_xref="SWISS-PROT:Q9CDF4"
/translation="MIESNESYSYGGDTIIEPLGTLSGFDLVRALREARAAACAGKD
AGRWNVGPPVFPVTRDRNNWSPGPDVDRPQGLKVAHDLAKKRGSAQVAGRVFG
QWASNVGQIADHAFPPVGLNNWGSVLTAEATWATQIRIMQALLAKIAAAGVGVVT
SLKITGTPAPSWRKGPHIAGRGRDITYG"

```

**IMPORTANT:** This sequence is not the entire insert of clone RP11-456A18. It may be



FEATURES  
source  
Location/Qualifiers  
1..158559  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-257L6"  
/clone\_lib="RP11-11"

ORIGIN  
Query Match 70.7%; Score 21.2; DB 9; Length 158559;  
Best Local Similarity 88.5%; Pred. NO. 2e+02;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGGGCGAGAACCGAGGTAGACCAA 26  
Db 132869 TAGGCGAGAACCGAGGTAGACCAA 132894

RESULT 10  
AC069541/c  
LOCUS Homo sapiens chromosome 10 clone RP11-38902, complete sequence.  
DEFINITION Homo sapiens chromosome 10 clone RP11-38902, complete sequence.  
ACCESSION AC069541  
VERSION AC069541.5 GI:20330781  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 158568)  
AUTHORS Smith,D.R.  
TITLE Genome Therapeutics Corporation Sequencing Center: Human Genome  
SEQUENCE DATA  
Unpublished  
REFERENCE 2 (bases 1 to 158568)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
REFERENCE 3 (bases 1 to 158568)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (13-APR-2001) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
REFERENCE 4 (bases 1 to 158568)  
AUTHORS Smith,D.R.  
TITLE Direct Submission  
JOURNAL Submitted (26-APR-2002) Genome Therapeutics Corporation, 100 Beaver  
Street, Waltham, MA 02453, USA  
REMARK 1329 bases of Tn10 (J01829) transposon removed here  
COMMENT On Apr 26, 2002 this sequence version replaced gi:13621215.  
FEATURES  
source  
Location/Qualifiers  
1..158568  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-38902"  
/clone\_lib="RP11-11"

ORIGIN  
Query Match 70.7%; Score 21.2; DB 9; Length 158568;  
Best Local Similarity 88.5%; Pred. NO. 2e+02;  
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGGGCGAGAACCGAGGTAGACCAA 26  
Db 25700 TAGGCGAGAACCGAGGTAGACCAA 25675

RESULT 11  
AC118678/c  
LOCUS Mus musculus clone RP23-406A5, WORKING DRAFT SEQUENCE, 11 unordered  
pieces.  
DEFINITION Mus musculus (house mouse)  
ACCESSION AC118678  
VERSION AC118678.3 GI:29164651  
KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 188591)  
AUTHORS Birren,B., Nusbaum,C. and Lander,E.  
TITLE Mus musculus, clone RP23-406A5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 188591)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,  
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,  
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,  
Landers,T., Lehotzky,J., Levine,R., Lindblad-Toh,K., Liu,G.,  
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,  
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,  
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,  
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,  
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,  
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,  
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 188591)  
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,J.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
McCarthy,M., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (23-MAR-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Mar 23, 2003 this sequence version replaced gi:28269643.  
All repeats were identified using RepeatMasker:  
Smit,A.P.A. & Green,P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

```

----- Project Information
Center project name: L19876
Center clone name: 406_A_5
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 186196 bases at least Q40
Consensus quality: 186891 bases at least Q30
Consensus quality: 187169 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 187591; sum-of-contigs
Quality coverage: 9.7 in Q20 bases; agarose-fp
Quality coverage: 9.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5777: contig of 5777 bp in length
* 5778 5877: gap of 100 bp
* 5878 7818: contig of 1941 bp in length
* 7819 7919: gap of 100 bp
* 7919 11201: contig of 3283 bp in length
* 11202 11301: gap of 100 bp
* 11302 18882: contig of 7581 bp in length
* 18883 18982: gap of 100 bp
* 18983 27416: contig of 8434 bp in length
* 27417 27516: gap of 100 bp
* 27517 48540: contig of 21024 bp in length
* 48541 48641: gap of 100 bp
* 48641 66525: contig of 17885 bp in length
* 66526 66626: gap of 100 bp
* 66626 93177: contig of 26552 bp in length
* 93178 93278: gap of 100 bp
* 93278 114625: contig of 21348 bp in length
* 114626 114725: gap of 100 bp
* 114726 143496: contig of 28770 bp in length
* 143496 143595: gap of 100 bp
* 143595 188591: contig of 44396 bp in length.
*
* Location/Qualifiers
* 1..188591
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP23-406A5"
* /clone_lib="RPCI-23 Female Mouse BAC"
* 1..5777
* /note="assembly_fragment
* clone_end:SP6
* vector_side:left"
*
* 5878..7818
* /note="assembly_fragment"
* 7919..11201
* /note="assembly_fragment"
* 11302..18882
* /note="assembly_fragment"
* 18983..27416
* /note="assembly_fragment"
* 27517..48540
* /note="assembly_fragment"
* 48641..66525
* /note="assembly_fragment"
* 66626..93177
* /note="assembly_fragment"
* 93278..114625
* /note="assembly_fragment"
* 114726..143495
* /note="assembly_fragment"

```

## FEATURES Source

```

misc_feature 143596..188591
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN

Query Match 70.0%; Score 21; DB 2; Length 188591;
Best Local Similarity 82.8%; Pred. No. 2.6e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGCGCAGACCCAGGAGTAGCACCACCAATGAG 30
|||||
Db 135354 GGCAAAACCCAGGAGGACGACCAATGAG 135326

RESULT 12
AC105077/c
LOCUS AC105077 208342 bp DNA linear HTG 21-AUG-2002
DEFINITION Mus musculus clone RP23-390G1, WORKING DRAFT SEQUENCE, 8 ordered
pieces.
ACCESSION AC105077
VERSION AC105077.4 GI:22381833
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 208342)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-390G1
Unpublished
REFERENCE 2 (bases 1 to 208342)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Choepe,I., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Govette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K.,
Lamazares,R., Landers,T., Lenoczky,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 208342)
REFERENCE 3 (bases 1 to 208342)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,I., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,

```

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

## misc\_feature

Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
Zembek, L., Zimmer, A. and Zody, M.

**TITLE**  
JOURNAL  
**COMMENT**  
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 21, 2002 this sequence version replaced gi:19774548.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L19875

Center clone name: 390\_G\_1

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 205782 bases at least Q40

Consensus quality: 207270 bases at least Q30

Consensus quality: 207451 bases at least Q20

Insert size: 200000; agarose-fp

Insert size: 207642; sum-of-contigs

Quality coverage: 10.8 in Q20 bases; agarose-fp

Quality coverage: 10.4 in Q20 bases; sum-of-contigs

-----  
\* NOTE: this is a 'working draft' sequence. It currently  
\* consists of 8 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and

\* by the accession number will be preserved.

\* 1 10477: contig of 10477 bp in length

\* 10478 10577: gap of 100 bp

\* 10578 14116: contig of 3539 bp in length

\* 14117 14216: gap of 100 bp

\* 14217 17666: contig of 3450 bp in length

\* 17667 17767: gap of 100 bp

\* 17767 38791: contig of 21025 bp in length

\* 38792 38891: gap of 100 bp

\* 38892 59740: contig of 20849 bp in length

\* 59741 59840: gap of 100 bp

\* 59841 96618: contig of 36778 bp in length

\* 96619 96719: gap of 100 bp

\* 96719 149994: contig of 53276 bp in length

\* 149995 150094: gap of 100 bp

\* 150095 208342: contig of 58248 bp in length.

----- Location/Qualifiers

1. 208342

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="RP23-390G1"

/clone\_lib="RPCI-23 Female Mouse BAC"

1. 10477

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left

10578. 14116

/note="assembly\_fragment"

14217. 17666

/note="assembly\_fragment"

17767. 38791

/note="assembly\_fragment"

38892. 159740

/note="assembly\_fragment"

misc\_feature 59841..96618

/note="assembly\_fragment"

misc\_feature 96719..149994

/note="assembly\_fragment"

misc\_feature 150095..208342

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

#### ORIGIN

Query Match 70.0%; Score 21; DB 2; Length 208342;

Best Local Similarity 82.8%; Pred. No. 2.6e+02;

Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 GGCGCAGAACACGAGGATAGCACCACATGAG 30

||||| ||||||| ||||||| ||||||| |||||||

Db 5869 GGCACAAACACGAGGACACACATGAG 5841

#### RESULT 13

AR413807/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 510

/organism="unknown"

/mol\_type="genomic DNA"

#### ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 510;

Best Local Similarity 85.2%; Pred. No. 2.3e+02;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCGCGACAGACGAGGATAGCACCACAT 27

||||| ||||||| ||||||| ||||||| |||||||

Db 125 TGCGCGTGAACGAGGATAGCAGAGT 99

#### RESULT 14

BD109360/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

OS Homo sapiens (human)

PN JP 2002010789-A/1437

PD 15-JAN-2002

PF 07-AUG-2000 JP 2002080989

PR 05-AUG-1999 US 60/147499

PI JEAN BAPTIST DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI

GIORDANO

PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC

C12N1/21,

BD109360 510 bp DNA linear PAT 18-SEP-2002

EST and encoded human protein.

BD109360

BD109360.1 GI:23204178

JP 2002010789-A/1437

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 510)

Edwards, J.B.D.M., Jobert, S. and Giordano, J.E.

EST and encoded human protein

Patent: JP 2002010789-A 1437 15-JAN-2002;

GENSET CORP

```

PC C12N5/10,C12P21/02,C12P21/08,C12Q1/68,C12N15/00,C12N5/00, PC
C12N15/00
CC Von Heijne matrix
CC score 7.6999980926514
CC seq LLFLALLTGLLL/LV
FT Key Location/Qualifiers
FT CDS 58..363
FT sig_peptide 58..111.
FT sig Location/Qualifiers
FEATURES
    source
        1..510
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 510;
Best Local Similarity 85.2%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 125 TGGCGCTGACCCAGGAGTAGCAGACT 99

RESULT 15
E10616/c
LOCUS Human cDNA encoding cytochrome P4502B6.
DEFINITION
E10616
ACCESSION
E10616
VERSION JP 1996027193-A/3.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1476)
AUTHORS Funae,Y., Imaoka,S., Matsuki,Y., Hayashi,K. and Yabusaki,Y.
TITLE ANTIBODY RECOGNIZING CYTOCHROME P4502B6 ORIGINATED FROM MAN
JOURNAL Patent: JP 1996027193-A 3 30-JAN-1996;
SUMITOMO CHEM CO LTD
COMMENT
OS Homo sapiens (human)
PN JP 1996027193-A/3
PD 30-JAN-1996
PF 13-JUL-1994 JP 1994161548
PI FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
YABUSAKI YOSHIYASU
PC C07K16/18,C12N15/09//C12N1/19,C12N9/02,G01N33/53,(C12N1/19, PC
C12R1:865);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1476
FT FT /organism="Homo sapiens"
FT FT /tissue type="liver"
FT CDS 1..1476
FT FT /product="cytochrome P4502B6".
FT Location/Qualifiers
FEATURES
    source
        1..1476
            /organism="Homo sapiens"
            /mol_type="genomic RNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 1476;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 68 TGGCGCTGACCCAGGAGTAGCAGACT 42

PC C12N5/10,C12P21/02,C12P21/08,C12Q1/68,C12N15/00,C12N5/00, PC
C12N15/00
CC Von Heijne matrix
CC score 7.6999980926514
CC seq LLFLALLTGLLL/LV
FT Key Location/Qualifiers
FT CDS 58..363
FT sig_peptide 58..111.
FT sig Location/Qualifiers
FEATURES
    source
        1..510
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 510;
Best Local Similarity 85.2%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 125 TGGCGCTGACCCAGGAGTAGCAGACT 99

RESULT 15
E10616/c
LOCUS Human cDNA encoding cytochrome P4502B6.
DEFINITION
E10616
ACCESSION
E10616
VERSION JP 1996027193-A/3.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1476)
AUTHORS Funae,Y., Imaoka,S., Matsuki,Y., Hayashi,K. and Yabusaki,Y.
TITLE ANTIBODY RECOGNIZING CYTOCHROME P4502B6 ORIGINATED FROM MAN
JOURNAL Patent: JP 1996027193-A 3 30-JAN-1996;
SUMITOMO CHEM CO LTD
COMMENT
OS Homo sapiens (human)
PN JP 1996027193-A/3
PD 30-JAN-1996
PF 13-JUL-1994 JP 1994161548
PI FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
YABUSAKI YOSHIYASU
PC C07K16/18,C12N15/09//C12N1/19,C12N9/02,G01N33/53,(C12N1/19, PC
C12R1:865);
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1476
FT FT /organism="Homo sapiens"
FT FT /tissue type="liver"
FT CDS 1..1476
FT FT /product="cytochrome P4502B6".
FT Location/Qualifiers
FEATURES
    source
        1..1476
            /organism="Homo sapiens"
            /mol_type="genomic RNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 1476;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 68 TGGCGCTGACCCAGGAGTAGCAGACT 42

```

```

RESULT 16
E10861/c
LOCUS cDNA encoding human cytochrome P450.
DEFINITION
E10861
ACCESSION
E10861
VERSION JP 1996056695-A/10.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1476)
AUTHORS Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
Nakatsuka,I.
TITLE METHOD FOR EVALUATING SAFETY
JOURNAL Patent: JP 1996056695-A 10 05-MAR-1996;
SUMITOMO CHEM CO LTD
COMMENT
OS Homo sapiens (human)
PN JP 1996056695-A/10
PD 05-MAR-1996
PF 15-JUL-1994 JP 1994164184
PR 20-JUL-1993 JP 93P 201120, 30-JUL-1993 JP 93P 208279, PR
17-JUN-1994 JP 94P 136053
PI HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU, PI KOMAI
KOICHIRO,
KI KANEKO HIDEO, NAKATSUKA IWAO
PC C12Q1/02,C12M1/34,C12Q1/26;
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FH Key Location/Qualifiers
FT source 1..1476
FT CDS 1..1476
FT /product="human cytochrome P450 2B6".
FEATURES
    source
        1..1476
            /organism="Homo sapiens"
            /mol_type="genomic RNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match 68.7%; Score 20.6; DB 6; Length 1476;
Best Local Similarity 85.2%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27
Db 68 TGGCGCTGACCCAGGAGTAGCAGACT 42

RESULT 17
AR399329/c
LOCUS AR399329 1476 bp DNA linear PAT 18-DEC-2003
DEFINITION
AR399329 Sequence 10 from patent US 6620593.
ACCESSION
AR399329
VERSION AR399329.1 GI:40141131
KEYWORDS
SOURCE
ORGANISM Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 1476)
AUTHORS Hayashi,K., Sakaki,T., Yabusaki,Y., Komai,K., Kaneko,H. and
Nakatsuka,I.
TITLE Method for safety evaluation of chemical compound using recombinant
yeast expressing human cytochrome P450
JOURNAL Patent: US 6620593-A 10 16-SEP-2003;
FEATURES
    source
        1..1476
            /organism="unknown"
            /mol_type="genomic DNA"

```



[illegible]

ADAM VORN (GB); STRATFORD IAN JAMES (GB); GRIFFITHS LEIGH (GB); KAN  
ON (GB); OXFORD BIOMEDICA LTD (GB)

## FEATURES

source

Location/Qualifiers  
1..3045  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
|||||  
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48  
|||||

## RESULT 21

AX018627/c  
LOCUS 3045 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 45 from Patent WO9945126.  
ACCESSION AX018627  
VERSION AX018627.1 GI:10042755

## KEYWORDS

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Kingsman, S.M., Mitrophanous, K., Patterson, A.V., Stratford, I.J.,  
Griffiths, L. and Kan, O.  
TITLE Enhanced prodrug activation  
JOURNAL Patent: WO 9945126-A 45 10-SEP-1999;  
KINGSMAN SUSAN MARY (GB); MITROPHANOUS KYRIACOS (GB); PATTERSON  
ADAM VORN (GB); STRATFORD IAN JAMES (GB); GRIFFITHS LEIGH (GB); KAN  
ON (GB); OXFORD BIOMEDICA LTD (GB)

## FEATURES

source

1..3045  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
|||||  
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48  
|||||

## RESULT 22

AX332736/c  
LOCUS 3045 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 3245 from Patent WO0194629.  
ACCESSION AX332736  
VERSION AX332736.1 GI:18123370

## KEYWORDS

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
Horrigian, S., Soppet, D.R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
JOURNAL gene sets  
Patent: WO 0194629-A 3245 13-DEC-2001;  
Avalon Pharmaceuticals (US)

## FEATURES

source

1..3045

/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
|||||  
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48  
|||||

## RESULT 23

AX409672/c  
LOCUS 3045 bp DNA linear PAT 14-JUN-2002  
DEFINITION Sequence 2319 from Patent WO0229103.  
ACCESSION AX409672  
VERSION AX409672.1 GI:21442377

## KEYWORDS

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.  
TITLE Gene expression profiles in liver cancer  
JOURNAL Patent: WO 0229103-A 2319 11-APR-2002;  
GENE LOGIC INC (US)

## FEATURES

source

1..3045  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="EMBL/GenBank Accession No. M29874"

## ORIGIN

Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
|||||  
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48  
|||||

## RESULT 24

BD136285/c  
LOCUS 3045 bp DNA linear PAT 18-SEP-2002  
DEFINITION Enhancement of prodrug activation.  
ACCESSION BD136285  
VERSION BD136285.1 GI:23231230

## KEYWORDS

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS 1 (bases 1 to 3045)  
Stratford, I.J., Patterson, A.V., Kingsman, S.M., Kan, O., Griffiths, L.  
and Mitrophanous, K.  
TITLE Enhancement of prodrug activation  
JOURNAL Patent: JP 2002505341-A 38 19-FEB-2002;  
OXFORD BIOMEDICA LTD

## COMMENT

OS Homo sapiens (human)  
PN JP 2002505341-A/38  
PD 19-FEB-2002  
PF 05-MAR-1999 JP 2000534657  
PR 06-MAR-1998 GB 9804841.6, 19-AUG-1998 GB 9818103.5 PR  
29-JAN-1999 GB 9902081.0  
PI IAN JAMES STRATFORD, ADAM VORN PATTERSON, SUSAN MARY KINGSMAN, ON  
PI KAN,

PI LEIGH GRIFFITHS, KYRIACOS MITROPHANOUS  
PC A61K47/48,A61K35/76,A61K38/44,A61K45/00,A61K48/00,A61P9/10, PC  
A61P29/00,  
PC A61P35/00,A61P43/00,C12N5/10,C12N7/00,C12N9/02,C12N15/09, PC  
A61K37/50,  
PC C12N5/00,C12N15/00  
CC Enhancement of prodrug activation  
FH Key Location/Qualifiers  
FT source 1..3045 /organism="Homo sapiens (human)".  
FT Location/Qualifiers  
1..3045  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

FEATURES  
source  
1..3045  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TGGCGCAGAACGAGTAGCACCACAT 27  
|||||  
Db 74 TGGCGCTGAACGAGTAGCAAGAGT 48  
|||||

RESULT 25  
HUMCYP2BB/c 3045 bp mRNA linear PRI 02-NOV-1994  
LOCUS Human cytochrome P450-IIB (h1B1) mRNA, complete cds.  
DEFINITION M29874 J02864  
ACCESSION M29874.1 GI:181295  
VERSION cytochrome P450; cytochrome P450 IIB.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3045)  
AUTHORS Yamano,S., Nambu,P.T., Aoyama,T., Meyer,U.A., Inaba,T.,  
Kallow,W., Gelboin,H.V., McBride,O.W. and Gonzalez,F.J.  
TITLE cDNA cloning and sequence and cDNA-directed expression of human  
P450 IIB1: identification of a normal and two variant cDNAs derived  
from the CYP2B locus on chromosome 19 and differential expression  
of the IIB mRNAs in human liver  
JOURNAL Biochemistry 28 (18), 7340-7348 (1989)  
MEDLINE 90057429  
PUBMED 2573390  
COMMENT Original source text: Human liver, cDNA to mRNA.  
FEATURES  
source  
1..3045  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/map="19q13.2"  
1..3045  
/gene="CYP2B"  
7..1482  
/gene="CYP2B"  
/note="cytochrome P450-IIB"  
/codon\_start=1  
/protein\_id="AAA52144.1"  
/db\_xref="GI:181296"  
/db\_xref="GDB:G00-120-752"  
/translation="MELSVLLFLALLTGLLLLVQRPNTDRLPPGPRLPLLGNLL  
QMDRGLKFLFRKYGDFVTHLGPVVMVLCGVEAIRALVDKAEAFSGRGRKIA  
KVDPFFRGYVIFANGNRKVLRRFSVTTMRDFGMRKSVVEERIOBEACGLIELRKS  
KGLMDPTFTANIIQSVFGKRFHQDFGLKMLNLFQTFSLISSVFGQFPE  
LFSGLFVYFPMHROVYNLQELINAYIGHVSKHRETLDPSPAKDLIDTVLLHWEKEK  
SNAHSEFHNQANLNTSLFACQETSTTLRYGELLMLKYPHVARVREIQVIGP  
HRPELHDKRMKPYTEAVIYEQPSDLEPMGVPHVQTSTSRGIIIPNDIEVFLIL  
STALHDHYEKPDAFNPDLFDANGALKKTEAIPFSLGKRICLGEIARAEFLFF  
TTLQNFSPMASPVAPEDIDLTPOEGVGKIPPTYQIRFLPR"  
1716..1997

repeat\_region

repeat\_region /note="Alu repeat 1"  
2134..2306  
/note="Alu repeat 2"

ORIGIN  
Query Match 68.7%; Score 20.6; DB 9; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TGGCGCAGAACGAGTAGCACCACAT 27  
|||||  
Db 74 TGGCGCTGAACGAGTAGCAAGAGT 48  
|||||

RESULT 26  
AC011541/c 39944 bp DNA linear PRI 20-JUL-2001  
LOCUS Homo sapiens chromosome 19 clone LLNLR-258F7, complete sequence.  
DEFINITION AC011541  
ACCESSION AC011541  
VERSION AC011541.7 GI:14971181  
KEYWORDS HTG; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 39944)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
2 (bases 1 to 39944)  
REFERENCE DOE Joint Genome Institute.  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 39944)  
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (20-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jul 20, 2001 this sequence version replaced gi:14042992.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence;  
Estimated Total Number of Errors is 0.  
FEATURES  
source  
1..39944  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="LLNLR-258F7"

ORIGIN  
Query Match 68.7%; Score 20.6; DB 9; Length 39944;  
Best Local Similarity 85.2%; Pred. No. 3.4e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TGGCGCAGAACGAGTAGCACCACAT 27  
|||||  
Db 37747 TGGCGCTGAACGAGTAGCAAGAGT 37721  
|||||

RESULT 27  
AC023172/c 42547 bp DNA linear PRI 09-FEB-2000  
LOCUS Homo sapiens chromosome 19, cosmid F22376, complete sequence.  
DEFINITION AC023172  
ACCESSION AC023172  
VERSION AC023172.1 GI:6957691  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens



repeat\_region 18249..18321  
/rpt family=" (CAGAGA)n"  
repeat\_region 18322..18594  
/rpt family="AluSc"  
repeat\_region 18595..18619  
/rpt family=" (TAA)n"  
repeat\_region 18624..18720  
/rpt family=" (CAGAGA)n"  
repeat\_region complement(18961..19259)  
/rpt family="AluJo"  
repeat\_region 19395..19577  
/rpt family="MIR"  
repeat\_region 20033..20221  
/rpt family="GA-rich"  
repeat\_region complement(20268..20434)  
/rpt family="AluSc"  
repeat\_region complement(20435..20733)  
/rpt family="AluSc"  
repeat\_region complement(21020..21241)  
/rpt family="AluSg/x"  
repeat\_region complement(21270..21432)  
/rpt family="MERSA"  
repeat\_region 21675..21855  
/rpt family="GA-rich"  
repeat\_region 22215..22440  
/rpt family="G-rich"  
repeat\_region 22438..22646  
/rpt family=" (CAGAGA)n"  
repeat\_region complement(23201..23211)  
/rpt family="L1PB3"  
repeat\_region complement(23212..23463)  
/rpt family="AluY"  
repeat\_region complement(23464..23720)  
/rpt family="L1PB3"  
repeat\_region 23736..23851  
/rpt family=" (TTCA)n"  
repeat\_region 23911..23958  
/rpt family=" (TTCA)n"  
repeat\_region 24033..24325  
/rpt family=" (TTCA)n"  
repeat\_region 24326..24351  
/rpt family=" (ATTG)n"  
repeat\_region 24352..24384  
/rpt family=" (TTCA)n"  
repeat\_region 24543..24568  
/rpt family=" (TG)n"  
repeat\_region complement(24763..25059)  
Query Match 68.7%; Score 20.6; DB 9; Length 42547;  
Best Local Similarity 85.2%; Pred. No. 3.4e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TGGCGCAGACGAGGTAGCACCAGT 27  
DB 4183 TGGCGCTGACCGAGTAGCAGAGT 4157  
RESULT 28  
AC092374/c AC092374 150962 bp DNA linear PRI 28-FEB-2003  
LOCUS Homo sapiens chromosome 16 clone RP11-63716, complete sequence.  
DEFINITION AC092374 AC022789  
ACCESSION AC092374.4 GI:28603936  
VERSION HTG.  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 150962)  
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los  
Alamos National Laboratory.  
TITLE Direct Submission  
JOURNAL Unpublished  
COMMENT

REFERENCE 2 (bases 1 to 150962)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 150962)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (27-MAR-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 4 (bases 1 to 150962)  
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los  
Alamos National Laboratory.  
TITLE Direct Submission  
JOURNAL Submitted (28-FEB-2003) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Feb 28, 2003 this sequence version replaced gi:19745005.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center and Los Alamos  
National Laboratory  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
FEATURES  
Location/Qualifiers  
source 1..150962  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosomes="16"  
/clone="RP11-63718"  
ORIGIN  
Query Match 68.7%; Score 20.6; DB 9; Length 150962;  
Best Local Similarity 85.2%; Pred. No. 3.8e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 GGGCGAGAACGAGGTAGCACCAGT 28  
DB 128079 GGGCGAARACGAGGTAAACCAATG 128053  
RESULT 29  
AC145155/c AC145155 164831 bp DNA linear HTG 10-JUN-2003  
LOCUS Homo sapiens chromosome 5 clone RP11-99N22, WORKING DRAFT SEQUENCE.  
DEFINITION AC145155 GI:31560162  
ACCESSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
VERSION AC145155.1  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 164831)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
COMMENT  
REFERENCE 2 (bases 1 to 164831)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUN-2003) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 444715  
Center clone name: RPCI-11\_99N22  
-----  
Summary Statistics  
Consensus quality: 164573 bases at least Q40

Consensus quality: 164754 bases at least Q30  
 Consensus quality: 164829 bases at least Q20  
 Estimated insert size: 175000; agarose-fp estimation  
 Estimated insert size: 164831; sum-of-contigs estimation  
 Quality coverage: 11.07 in Q20 bases; agarose-fp estimation  
 Quality coverage: 11.75 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 164831: contig of 164831 bp in length.

## FEATURES

Location/Qualifiers  
 1. .164831  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="RP11-99N22"  
 /clone\_lib="RP11-99N22"  
 /clone\_lib="RP11-99N22"

## ORIGIN

Query Match 58.7%; Score 20.6; DB 2; Length 164831;  
 Best Local Similarity 85.2%; Pred. No. 3.9e-02;  
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCGCAGACACAGGAGTAGCACCACATG 28

Db 12235 GCGCAGACACAGGAGTAGCACCACATG 12209

## RESULT 30

AC117614/c 222809 bp DNA linear RCD 01-NOV-2003  
 Mus musculus chromosome 5, clone RP23-110C17, complete sequence.

## DEFINITION

AC117614

AC117614.12 GI:38142550

HTG.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 222809)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus chromosome 5, Clone RP23-110C17

Unpublished

2 (bases 1 to 222809)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,

Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,

Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A.,

Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,

Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,

Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,

Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,

Landers,T., Lehotzky,J., Levine,R., Lindblad-Toh,K., Liu,G.,

MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,

Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,

Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,

Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,

Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,

Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,

Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,

Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,

Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## TITLE

Direct Submission

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 222809)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (11-OCT-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 222809)  
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (01-NOV-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Nov 1, 2003 this sequence version replaced gi:37651805.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L23357  
 Center clone name: 110\_C\_17  
 ----- Location/Qualifiers  
 1. .222809  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="5"  
 /map="5"  
 /clone="RP23-110C17"  
 /clone\_lib="RP11-99N22"  
 /clone\_lib="RP11-99N22"

## FEATURES

## source

1. .222809

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="5"

/map="5"

/clone="RP23-110C17"

/clone\_lib="RP11-99N22"

/clone\_lib="RP11-99N22"

1. .11631

misc\_feature

1. .11631

misc\_feature

1. .11631

misc\_feature

1. .11631

misc\_feature

1. .11631

```
/note="wgs_end_extension
clone_end:SP6"
repeat_region 1648..1729
repeat_region /rpt_family="MTE"
repeat_region 2408..2619
repeat_region /rpt_family="B3"
repeat_region 3221..3241
repeat_region /rpt_family="AT rich"
repeat_region 3787..3816
repeat_region /rpt_family="(TGG)n"
repeat_region 4265..4344
repeat_region /rpt_family="(TCCA)n"
repeat_region complement(5866..5984)
repeat_region /rpt_family="ID_B1"
repeat_region 6022..6093
repeat_region /rpt_family="B1_MM"
repeat_region 6181..6309
repeat_region /rpt_family="B1_MM"
repeat_region 6310..6363
repeat_region /rpt_family="(CAAAAA)n"
repeat_region 6557..6950
repeat_region /rpt_family="MTD"
repeat_region 6963..7144
repeat_region /rpt_family="LIMB7"
repeat_region 7150..7357
repeat_region /rpt_family="B3"
repeat_region complement(7378..7524)
repeat_region /rpt_family="B1_MM"
repeat_region 7552..7677
repeat_region /rpt_family="RSINE1"
repeat_region 7692..7721
repeat_region /rpt_family="(CAAT)n"
repeat_region 7733..7948
repeat_region /rpt_family="LIMB7"
repeat_region complement(8580..9057)
repeat_region /rpt_family="IAPLTR2_MM"
repeat_region 10093..10284
repeat_region /rpt_family="B3"
repeat_region 10291..10402
repeat_region /rpt_family="PB1D9"
repeat_region 10415..10451
repeat_region /note="x30 qual SNGL region"
repeat_region 11112..11213
repeat_region /rpt_family="PB1D7"
repeat_region 11215..11259
repeat_region /rpt_family="GA-rich"
repeat_region 11265..11382
repeat_region /rpt_family="B1_MM"
repeat_region 11383..11413
repeat_region /rpt_family="(CAAA)n"
repeat_region 11427..11976
repeat_region /rpt_family="LIM4"
repeat_region 11532..11637
repeat_region /note="clone_boundary
clone_end:SP6
site:ECORI"
repeat_region 13247..13399
repeat_region /rpt_family="MER2"
repeat_region 13662..13761
repeat_region /rpt_family="PB1D7"
repeat_region 14295..14411
repeat_region /rpt_family="B1_MM"
repeat_region 14992..15035
repeat_region /rpt_family="B2_MM2"
repeat_region 15023..15154
repeat_region /rpt_family="B3"
repeat_region 15528..15557
repeat_region /rpt_family="(GGA)n"
repeat_region 16277..16299
repeat_region /rpt_family="AT rich"
repeat_region complement(16384..16505)
repeat_region /rpt_family="B3A"
repeat_region 16593..16770
```

```
repeat_region /rpt_family="B2_MM2"
repeat_region 16799..16918
repeat_region /rpt_family="B2_MM2"
repeat_region complement(17278..17316)
repeat_region /rpt_family="MER5B"
repeat_region 17385..17487
repeat_region /rpt_family="B1_MM"
repeat_region 17686..17792
repeat_region /rpt_family="B1_MM"
repeat_region complement(17783..17847)
repeat_region /rpt_family="PB1D10"
repeat_region 18153..18172
repeat_region /rpt_family="(TTTG)n"
repeat_region complement(18173..18351)
repeat_region /rpt_family="B2_MM1"
repeat_region complement(19078..19471)
repeat_region /rpt_family="MTA"

Query Match 68.7%; Score 20.6; DB 10; Length 222809;
Best Local Similarity 85.2%; Pred.No.4e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCCCAAT 27
DB 118038 TGGCTCAGAACCATGAGTAACCCCAAT 118012
|||||
|||||
```

Search completed: June 20, 2004, 11:44:43  
Job time : 836.603 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 08:46:26 ; Search time 172.381 Seconds  
(without alignments)  
739.327 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcgacagaccagtagtagcaccatgag 30

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

N\_Geneseq 25Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result NO. | Score | Query Match % | Length | ID          | Description         |
|------------|-------|---------------|--------|-------------|---------------------|
| 1          | 30    | 100.0         | 1584   | AAH51956    | Aah51956 Mycobacte  |
| 2          | 30    | 100.0         | 110000 | AAI99682    | Aai99682 Mycobacte  |
| 3          | 30    | 100.0         | 110000 | AAI99683_00 | Aai99683 Mycobacte  |
| 4          | 20.6  | 68.7          | 1476   | AAQ87723_00 | Aac87723 Human aux  |
| 5          | 20.6  | 68.7          | 1476   | AAI28389    | Aac28389 Human cyt  |
| 6          | 20.6  | 68.7          | 1476   | AAI17417    | Aac17417 Human der  |
| 7          | 20.6  | 68.7          | 2915   | AAI11662    | Aas11662 Human CYP  |
| 8          | 20.6  | 68.7          | 3045   | AAZ19787    | Aaz19787 Human cyt  |
| 9          | 20.6  | 68.7          | 3045   | AAZ207812   | Aaz207812 Human cyt |
| 10         | 20.6  | 68.7          | 3045   | ABU64908    | Abi64908 Lung canc  |
| 11         | 20.6  | 68.7          | 3045   | ABT10310    | Abt10310 Human bre  |
| 12         | 20.6  | 68.7          | 3045   | ABN95821    | Abn95821 Gene #231  |
| 13         | 20.6  | 68.7          | 4887   | ADD29818    | Acc29818 Human tum  |
| 14         | 20.6  | 68.0          | 41     | AAI11651    | Aas11651 Human CYP  |
| 15         | 19.6  | 65.3          | 590    | AAH69361    | Aah69361 Human cer  |
| 16         | 19.6  | 65.3          | 649    | AAH72227    | Aah72227 Human cer  |
| 17         | 19.6  | 65.3          | 778    | AAH71164    | Aah71164 Human cer  |
| 18         | 19.6  | 65.3          | 2420   | AAH72859    | Aah72859 Human cer  |
| 19         | 19.6  | 65.3          | 2420   | ABV25077    | Abv25077 Human pro  |
| 20         | 19.6  | 65.3          | 2427   | ABL27011    | Abi27011 Drosophil  |
| 21         | 19.6  | 65.3          | 6200   | ABL27010    | Abi27010 Drosophil  |
| 22         | 19.4  | 64.7          | 537    | ABA61254    | Abi61254 Human foe  |
| 23         | 19.4  | 64.7          | 537    | AAI41155    | Aai41155 Probe #98  |

|    |      |      |        |   |             |                    |
|----|------|------|--------|---|-------------|--------------------|
| 24 | 19.4 | 64.7 | 537    | 4 | ABA29093    | Abi29093 Probe #75 |
| 25 | 19.4 | 64.7 | 537    | 4 | AAK35442    | Aak35442 Human bon |
| 26 | 19.4 | 64.7 | 537    | 4 | AAK09550    | Aak09550 Human bra |
| 27 | 19.4 | 64.7 | 537    | 4 | ABS35177    | Abs35177 Human liv |
| 28 | 19.4 | 64.7 | 537    | 6 | ABS09801    | Abs09801 Human gen |
| 29 | 19.2 | 64.0 | 21706  | 2 | AAV52169    | Aav52169 Streptoco |
| 30 | 19.2 | 64.0 | 110000 | 7 | ABS56454_14 | Abi56454_14 o      |
| 31 | 19   | 63.3 | 280    | 3 | AAC16260    | Aac16260 Human sec |
| 32 | 19   | 63.3 | 393    | 8 | ACH17812    | Ach17812 Human adu |
| 33 | 19   | 63.3 | 637    | 6 | ABS72349    | Abs72349 Human gen |
| 34 | 19   | 63.3 | 843    | 6 | ABX97067    | Abx97067 Human NOV |
| 35 | 19   | 63.3 | 1174   | 7 | ABX05033    | Abx05033 Human nov |
| 36 | 19   | 63.3 | 40178  | 9 | ADE43862    | Ade43862 Human LIP |
| 37 | 19   | 63.3 | 40178  | 9 | ADE43863    | Ade43863 Polymorph |
| 38 | 19   | 63.3 | 95269  | 6 | ABQ67195_4  | Abi67195_4 o       |
| 39 | 19   | 63.3 | 110000 | 6 | ABQ69245_13 | Abi69245_13 o      |
| 40 | 18.8 | 62.7 | 216    | 4 | ABL27759    | Abi27759 Drosophil |
| 41 | 18.8 | 62.7 | 289    | 4 | AAI22848    | Aai22848 Human bre |
| 42 | 18.8 | 62.7 | 328    | 4 | AAI26095    | Aai26095 Human bre |
| 43 | 18.8 | 62.7 | 328    | 4 | AAI18165    | Aai18165 Human bre |
| 44 | 18.8 | 62.7 | 365    | 4 | AAI22519    | Aai22519 Human bre |
| 45 | 18.8 | 62.7 | 365    | 4 | AAI22557    | Aai22557 Human bre |
| 46 | 18.8 | 62.7 | 366    | 4 | AAI13845    | Aai13845 Human bre |
| 47 | 18.8 | 62.7 | 369    | 4 | AAI22588    | Aai22588 Human bre |
| 48 | 18.8 | 62.7 | 372    | 4 | AAI22711    | Aai22711 Human bre |
| 49 | 18.8 | 62.7 | 377    | 4 | AAI13986    | Aai13986 Human bre |
| 50 | 18.8 | 62.7 | 381    | 4 | AAI13650    | Aai13650 Human bre |
| 51 | 18.8 | 62.7 | 438    | 4 | AAI13688    | Aai13688 Human bre |
| 52 | 18.8 | 62.7 | 439    | 4 | AAI13758    | Aai13758 Human bre |
| 53 | 18.8 | 62.7 | 575    | 4 | AAI22625    | Aai22625 Human bre |
| 54 | 18.8 | 62.7 | 1015   | 3 | AAC59665    | Aac59665 Human sec |
| 55 | 18.8 | 62.7 | 1069   | 7 | AAI51043    | Aai51043 Human DME |
| 56 | 18.8 | 62.7 | 1143   | 2 | AAV42730    | Aav42730 Human cyt |
| 57 | 18.8 | 62.7 | 1406   | 3 | AAC59652    | Aac59652 Human sec |
| 58 | 18.8 | 62.7 | 1440   | 7 | ABX53820    | Abx53820 Human CDN |
| 59 | 18.8 | 62.7 | 1500   | 6 | ABQ76539    | Abq76539 C. albica |
| 60 | 18.8 | 62.7 | 1824   | 2 | AAI89089    | Aai89089 Avian 3-h |
| 61 | 18.8 | 62.7 | 2216   | 4 | ABL27758    | Abi27758 Drosophil |
| 62 | 18.8 | 62.7 | 2234   | 4 | ABL24922    | Abi24922 Drosophil |
| 63 | 18.8 | 62.7 | 2815   | 4 | ABL09420    | Abi09420 Drosophil |
| 64 | 18.8 | 62.7 | 2907   | 6 | ABL55916    | Abi55916 Lung canc |
| 65 | 18.8 | 62.7 | 2907   | 6 | ABL55008    | Abi55008 Lung canc |
| 66 | 18.8 | 62.7 | 2907   | 6 | ABL55513    | Abi55513 Lung canc |
| 67 | 18.8 | 62.7 | 2907   | 6 | ABN95820    | Abn95820 Gene #231 |
| 68 | 18.8 | 62.7 | 3704   | 4 | ABL09414    | Abi09414 Drosophil |
| 69 | 18.8 | 62.7 | 4321   | 4 | ABL18634    | Abi18634 Drosophil |
| 70 | 18.6 | 62.0 | 290    | 7 | ABX88273    | Abx88273 Corn ear- |
| 71 | 18.6 | 62.0 | 936    | 9 | ADC23952    | Adc23952 DNA seque |
| 72 | 18.6 | 62.0 | 1283   | 2 | AAI60284    | Aai60284 Nucleic a |
| 73 | 18.6 | 62.0 | 1406   | 4 | ABL06013    | Abi06013 Drosophil |
| 74 | 18.6 | 62.0 | 3184   | 4 | ABL18159    | Abi18159 Drosophil |
| 75 | 18.6 | 62.0 | 3480   | 4 | ABL06012    | Abi06012 Drosophil |
| 76 | 18.6 | 62.0 | 7138   | 4 | ABL18158    | Abi18158 Drosophil |
| 77 | 18.4 | 61.3 | 228    | 4 | ABL25185    | Abi25185 Drosophil |
| 78 | 18.4 | 61.3 | 333    | 3 | AAC57133    | Aac57133 Pinus rad |
| 79 | 18.4 | 61.3 | 394    | 3 | AAC07594    | Aac07594 Human sec |
| 80 | 18.4 | 61.3 | 546    | 6 | ABK55262    | Abk55262 Human col |
| 81 | 18.4 | 61.3 | 627    | 4 | ABL10291    | Abi10291 Drosophil |
| 82 | 18.4 | 61.3 | 729    | 4 | AAK89066    | Aak89066 Human dig |
| 83 | 18.4 | 61.3 | 729    | 4 | AAI57623    | Aai57623 Human col |
| 84 | 18.4 | 61.3 | 729    | 6 | ABS99800    | Abs99800 Genomic D |
| 85 | 18.4 | 61.3 | 729    | 9 | ABD92953    | Abd92953 Human col |
| 86 | 18.4 | 61.3 | 752    | 6 | ABQ56190    | Abq56190 Human ova |
| 87 | 18.4 | 61.3 | 778    | 4 | AAH33523    | Aah33523 Human col |
| 88 | 18.4 | 61.3 | 864    | 2 | AAI13704    | Aai13704 Enterococ |
| 89 | 18.4 | 61.3 | 864    | 6 | ABS99499    | Abs99499 Enterococ |
| 90 | 18.4 | 61.3 | 888    | 4 | AAK89065    | Aak89065 Human dig |
| 91 | 18.4 | 61.3 | 888    | 4 | AAI57622    | Aai57622 Human col |
| 92 | 18.4 | 61.3 | 888    | 6 | ABS99799    | Abs99799 Genomic D |
| 93 | 18.4 | 61.3 | 888    | 9 | ABD92952    | Abd92952 Human col |
| 94 | 18.4 | 61.3 | 910    | 4 | AAI92558    | Aai92558 Human pol |
| 95 | 18.4 | 61.3 | 931    | 5 | AAH23841    | Aah23841 Human tra |
| 96 | 18.4 | 61.3 | 1059   | 6 | ABZ14011    | Abz14011 Arabidops |

|     |      |      |        |   |           |              |       |      |      |        |   |           |           |              |
|-----|------|------|--------|---|-----------|--------------|-------|------|------|--------|---|-----------|-----------|--------------|
| 97  | 18.4 | 61.3 | 1059   | 7 | ADA68115  | Arabidops    | 170   | 18   | 60.0 | 1203   | 7 | ADA69469  | Ada59469  | Rice gene    |
| 98  | 18.4 | 61.3 | 1246   | 4 | AD05827   | Arabidops    | 171   | 18   | 60.0 | 1267   | 6 | ABN98257  | Abn98257  | Arabidops    |
| 99  | 18.4 | 61.3 | 1246   | 5 | AAD06666  | A. thalia    | c 172 | 18   | 60.0 | 1337   | 3 | ABC48518  | Abc48518  | Arabidops    |
| 100 | 18.4 | 61.3 | 1246   | 6 | ABK65295  | Arabidops    | c 173 | 18   | 60.0 | 1355   | 3 | AAC48403  | Aac48403  | Arabidops    |
| 101 | 18.4 | 61.3 | 1246   | 8 | ADA15568  | DNA encod    | c 174 | 18   | 60.0 | 1377   | 3 | AAC41116  | Aac41116  | Arabidops    |
| 102 | 18.4 | 61.3 | 1246   | 9 | ADB31774  | DNA encod    | 175   | 18   | 60.0 | 1447   | 3 | AAC41073  | Aac41073  | Arabidops    |
| 103 | 18.4 | 61.3 | 1246   | 9 | ADD55813  | Thalecres    | c 176 | 18   | 60.0 | 1452   | 3 | AAC41073  | Aac41073  | Arabidops    |
| 104 | 18.4 | 61.3 | 1246   | 9 | ADP30355  | Plant yle    | 177   | 18   | 60.0 | 1512   | 8 | ADB09996  | Adb09996  | Allolococ    |
| 105 | 18.4 | 61.3 | 1311   | 2 | AAC34833  | Arabidops    | 178   | 18   | 60.0 | 1512   | 8 | ADB09998  | Adb09998  | Allolococ    |
| 106 | 18.4 | 61.3 | 1373   | 2 | AAV31355  | Trichoder    | 179   | 18   | 60.0 | 1932   | 7 | ACD01710  | AcD01710  | DNA sequ     |
| 107 | 18.4 | 61.3 | 1384   | 3 | AAC45742  | Arabidops    | 180   | 18   | 60.0 | 1932   | 7 | ACD01715  | AcD01715  | DNA sequ     |
| 108 | 18.4 | 61.3 | 2100   | 2 | AAV31356  | Trichoder    | 181   | 18   | 60.0 | 1932   | 7 | ACD01722  | AcD01722  | DNA sequ     |
| 109 | 18.4 | 61.3 | 2288   | 4 | AB125164  | Drosophi     | 182   | 18   | 60.0 | 1932   | 7 | ACD01682  | AcD01682  | DNA sequ     |
| 110 | 18.4 | 61.3 | 2471   | 2 | AAK35516  | CDNA enco    | 183   | 18   | 60.0 | 1932   | 7 | ACD01689  | AcD01689  | DNA sequ     |
| 111 | 18.4 | 61.3 | 2855   | 6 | ABK35764  | CDNA sequ    | 184   | 18   | 60.0 | 1932   | 7 | ACD01713  | AcD01713  | DNA sequ     |
| 112 | 18.4 | 61.3 | 3443   | 4 | AB110290  | Drosophi     | 185   | 18   | 60.0 | 1932   | 7 | ACD01724  | AcD01724  | DNA sequ     |
| 113 | 18.4 | 61.3 | 3859   | 7 | ABT16586  | Ethylene     | 186   | 18   | 60.0 | 1932   | 9 | ADB71121  | AdB71121  | Adeno-ass    |
| 114 | 18.4 | 61.3 | 4581   | 5 | AAS69032  | DNA encod    | 187   | 18   | 60.0 | 1932   | 9 | ADB71086  | AdB71086  | Adeno-ass    |
| 115 | 18.4 | 61.3 | 7058   | 7 | AB224705  | Human cel    | 188   | 18   | 60.0 | 1932   | 9 | ADB71119  | AdB71119  | Adeno-ass    |
| 116 | 18.4 | 61.3 | 7931   | 4 | AAK85096  | Human imm    | 189   | 18   | 60.0 | 1932   | 9 | ADB71107  | AdB71107  | Adeno-ass    |
| 117 | 18.4 | 61.3 | 9731   | 4 | AAK72933  | Human imm    | 190   | 18   | 60.0 | 1932   | 9 | ADB71110  | AdB71110  | Adeno-ass    |
| 118 | 18.4 | 61.3 | 9733   | 4 | AAK72933  | Human imm    | 191   | 18   | 60.0 | 1932   | 9 | ADB71079  | AdB71079  | Adeno-ass    |
| 119 | 18.4 | 61.3 | 9733   | 4 | AAK85097  | Human imm    | 192   | 18   | 60.0 | 2166   | 6 | ABZ35214  | Abz35214  | Human gen    |
| 120 | 18.4 | 61.3 | 28011  | 4 | AAK81639  | Human imm    | 193   | 18   | 60.0 | 2479   | 4 | ABL06675  | AbL06675  | Drosophi     |
| 121 | 18.4 | 61.3 | 518    | 3 | AAC04044  | Human sec    | 194   | 18   | 60.0 | 3879   | 2 | AAV29009  | Aav29009  | Human pol    |
| 122 | 18.2 | 60.7 | 894    | 3 | AAK39565  | M. tuberc    | 195   | 18   | 60.0 | 3879   | 2 | AAV21060  | Aav21060  | CDNA enco    |
| 123 | 18.2 | 60.7 | 1239   | 9 | ACF69207  | Phototrab    | 196   | 18   | 60.0 | 4055   | 4 | AAK52394  | Aak52394  | Human pol    |
| 124 | 18.2 | 60.7 | 2306   | 9 | ADP45465  | Human gen    | c 197 | 18   | 60.0 | 4196   | 4 | ABL23927  | AbL23927  | Drosophi     |
| 125 | 18.2 | 60.7 | 5306   | 9 | ADP57302  | Human gen    | 198   | 18   | 60.0 | 5378   | 4 | ABL06674  | AbL06674  | Drosophi     |
| 126 | 18.2 | 60.7 | 5424   | 5 | AAK73915  | DNA encod    | 199   | 18   | 60.0 | 5378   | 4 | ABL23926  | AbL23926  | Drosophi     |
| 127 | 18.2 | 60.7 | 5956   | 5 | AAK73918  | DNA encod    | 200   | 18   | 60.0 | 6875   | 4 | ABL23926  | AbL23926  | Drosophi     |
| 128 | 18.2 | 60.7 | 7643   | 6 | ABQ99421  | Human cod    | 201   | 18   | 60.0 | 110000 | 2 | AAT42063  | Aat42063  | Continuation |
| 129 | 18.2 | 60.7 | 7931   | 9 | ADEB71077 | Human cod    | 202   | 18   | 60.0 | 110000 | 2 | AAT42063  | Aat42063  | Continuation |
| 130 | 18.2 | 60.7 | 10620  | 4 | AB117337  | Drosophi     | c 203 | 18   | 60.0 | 18998  | 6 | ABN85584  | Abn85584  | Human Her    |
| 131 | 18.2 | 60.7 | 12748  | 4 | AB117336  | Drosophi     | c 204 | 18   | 60.0 | 197496 | 6 | ABN85584  | Abn85584  | Human Her    |
| 132 | 18.2 | 60.7 | 110000 | 4 | AAI99682  | Continuation | c 205 | 17.8 | 59.3 | 61     | 9 | ADD68651  | Add68651  | DNA ampli    |
| 133 | 18.2 | 60.7 | 110000 | 4 | AAI99683  | Continuation | c 206 | 17.8 | 59.3 | 84     | 3 | AAZ89638  | Aaz89638  | HIV-2 ROD    |
| 134 | 18.2 | 60.7 | 110000 | 7 | ACF67367  | Continuation | c 207 | 17.8 | 59.3 | 84     | 3 | ADA07994  | Ada07994  | DNA encod    |
| 135 | 18.2 | 60.7 | 243072 | 7 | ACF65382  | Phototrab    | c 208 | 17.8 | 59.3 | 201    | 9 | ADD68664  | Add68664  | DNA ampli    |
| 136 | 18   | 60.0 | 406    | 6 | ABK27687  | Human col    | c 209 | 17.8 | 59.3 | 207    | 6 | ADD68664  | Add68664  | DNA ampli    |
| 137 | 18   | 60.0 | 413    | 7 | ABX42725  | Bovine ES    | 210   | 17.8 | 59.3 | 285    | 6 | ABL27250  | AbL27250  | Corn tass    |
| 138 | 18   | 60.0 | 424    | 7 | ABX35143  | Bovine ES    | 211   | 17.8 | 59.3 | 290    | 6 | ABL71455  | AbL71455  | Corn tass    |
| 139 | 18   | 60.0 | 427    | 4 | AA110161  | Probe #94    | 212   | 17.8 | 59.3 | 312    | 2 | AAT45452  | Aat45452  | Human LST    |
| 140 | 18   | 60.0 | 427    | 4 | ABK51793  | Human foe    | 213   | 17.8 | 59.3 | 321    | 7 | ABX41191  | Abx41191  | Bovine ES    |
| 141 | 18   | 60.0 | 427    | 4 | AAK31411  | Probe #97    | c 214 | 17.8 | 59.3 | 344    | 4 | AAI10501  | Aai10501  | Human bre    |
| 142 | 18   | 60.0 | 427    | 4 | ABK21623  | Probe #99    | c 215 | 17.8 | 59.3 | 353    | 7 | ABX35286  | Abx35286  | Bovine ES    |
| 143 | 18   | 60.0 | 427    | 4 | AAK25537  | Human bon    | c 216 | 17.8 | 59.3 | 392    | 4 | AAI85777  | Aai85777  | DNA encod    |
| 144 | 18   | 60.0 | 427    | 4 | AAK00102  | Human bra    | c 217 | 17.8 | 59.3 | 424    | 4 | AAI85777  | Aai85777  | DNA encod    |
| 145 | 18   | 60.0 | 427    | 4 | AAK00102  | Human bra    | c 218 | 17.8 | 59.3 | 435    | 6 | ABN96156  | Abn96156  | Gene #265    |
| 146 | 18   | 60.0 | 427    | 5 | AAI00102  | Probe #93    | 219   | 17.8 | 59.3 | 450    | 6 | AAZ894927 | Aaz894927 | Human DNA    |
| 147 | 18   | 60.0 | 427    | 6 | ABK00106  | Human gen    | c 220 | 17.8 | 59.3 | 466    | 3 | AAK00187  | Aac00187  | Human sec    |
| 148 | 18   | 60.0 | 427    | 2 | AAI19914  | Human gen    | c 221 | 17.8 | 59.3 | 550    | 4 | AAI21139  | Aai21139  | Human bre    |
| 149 | 18   | 60.0 | 468    | 4 | AAK53425  | Haemophil    | 222   | 17.8 | 59.3 | 565    | 3 | AAK07361  | Aac07361  | Human sec    |
| 150 | 18   | 60.0 | 468    | 7 | ACA34314  | Prokaryot    | 223   | 17.8 | 59.3 | 565    | 3 | AAK07362  | Aac07362  | Human sec    |
| 151 | 18   | 60.0 | 672    | 4 | AAI19454  | Probe #93    | 224   | 17.8 | 59.3 | 584    | 4 | AAH56802  | Aah56802  | Human imm    |
| 152 | 18   | 60.0 | 672    | 4 | ABK46469  | Human foe    | 225   | 17.8 | 59.3 | 635    | 6 | ABK83464  | Abk83464  | Human CDN    |
| 153 | 18   | 60.0 | 672    | 4 | AAI44646  | Probe #13    | 226   | 17.8 | 59.3 | 672    | 7 | ACC59937  | Acc59937  | Human IRA    |
| 154 | 18   | 60.0 | 672    | 4 | ABK46604  | Human bra    | 227   | 17.8 | 59.3 | 680    | 4 | AAI01553  | Aai01553  | Human sec    |
| 155 | 18   | 60.0 | 672    | 4 | ABK31607  | Probe #10    | c 228 | 17.8 | 59.3 | 707    | 9 | ADD68686  | Add68686  | DNA ampli    |
| 156 | 18   | 60.0 | 672    | 4 | ABK38653  | Human bon    | 229   | 17.8 | 59.3 | 720    | 4 | AAI22017  | Aai22017  | DNA encod    |
| 157 | 18   | 60.0 | 672    | 4 | AAK12928  | Human bra    | c 230 | 17.8 | 59.3 | 804    | 4 | AAI22066  | Aai22066  | DNA encod    |
| 158 | 18   | 60.0 | 672    | 4 | ABK38220  | Human liv    | 231   | 17.8 | 59.3 | 818    | 7 | ACC59953  | Acc59953  | Human IRA    |
| 159 | 18   | 60.0 | 672    | 5 | AAI05179  | Probe #51    | c 232 | 17.8 | 59.3 | 865    | 4 | AAI13196  | Aai13196  | Human bre    |
| 160 | 18   | 60.0 | 672    | 6 | ABK12723  | Human gen    | c 233 | 17.8 | 59.3 | 917    | 9 | ADC32514  | Adc32514  | Human nov    |
| 161 | 18   | 60.0 | 699    | 8 | ADB10002  | Allolococ    | 234   | 17.8 | 59.3 | 946    | 6 | AAH67870  | Aah67870  | C Glutami    |
| 162 | 18   | 60.0 | 704    | 7 | ABZ51337  | Aspergill    | c 235 | 17.8 | 59.3 | 1299   | 5 | AAH67870  | Aah67870  | C Glutami    |
| 163 | 18   | 60.0 | 1062   | 6 | ABN96948  | Arabidops    | c 236 | 17.8 | 59.3 | 1342   | 4 | AAI71533  | Aai71533  | Phototrab    |
| 164 | 18   | 60.0 | 1062   | 6 | ABZ14289  | Arabidops    | c 237 | 17.8 | 59.3 | 1386   | 7 | ACF67888  | Acf67888  | Phototrab    |
| 165 | 18   | 60.0 | 1136   | 6 | ABK63764  | Rat sequ     | 238   | 17.8 | 59.3 | 1489   | 6 | ABQ57598  | Abq57598  | Human MDD    |
| 166 | 18   | 60.0 | 1136   | 6 | ABT42020  | Toxicity     | 239   | 17.8 | 59.3 | 1599   | 4 | ADC30745  | Adc30745  | Human nov    |
| 167 | 18   | 60.0 | 1136   | 9 | ADB58357  | Toxicity     | 240   | 17.8 | 59.3 | 1657   | 9 | ABL10139  | AbL10139  | Drosophi     |
| 168 | 18   | 60.0 | 1136   | 9 | ADB58357  | Toxicity     | 241   | 17.8 | 59.3 | 1681   | 4 | ABL10139  | AbL10139  | Drosophi     |
| 169 | 18   | 60.0 | 1136   | 9 | ADB52928  | Primary r    | 242   | 17.8 | 59.3 | 1747   | 6 | ABZ25783  | Abz25783  | Human mul    |

Drug target; growth; organism viability; characterisation; ds.

Mycobacterium tuberculosis.

WO200135317-A1.

17-MAY-2001.

13-NOV-2000; 2000WO-US031152.

12-NOV-1999; 99US-0165086P.

12-NOV-1999; 99US-0165124P.

01-FEB-2000; 2000US-0179531P.

(REGC ) UNIV CALIFORNIA.

Eisenberg D, Rotstein SH, Marcotte EM;

WPI; 2001-329193/34.

P-PSDB; AAG81105.

Identifying nucleotide or polypeptide sequence for use as drug target,

Involves providing algorithm that analyzes a functional relationship

between nucleotide or polypeptide sequences, and comparing the sequences.

PS Disclosure; Page 53; 207pp; English.

This invention relates to a method for identifying a nucleotide or

polypeptide sequence that may be a drug target, or essential for growth

or viability of an organism. Polynucleotide sequences AAHS1947 - AAHS2092

represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium

tuberculosis proteins which are potential drug targets. The DNA and

protein sequences are used to illustrate the method of the invention. The

method involves providing an unknown nucleotide or polypeptide sequences,

and comparing it to a number of sequences along with at least one

algorithm capable of analysing a functional relationship between

nucleotide and polypeptide sequences. The method is useful for

characterising the function of nucleic acids and polypeptides that may be

useful as a target for a drug or essential for the growth or viability of

an organism

XX Sequence 1584 BP; 299 A; 539 C; 512 G; 234 T; 0 U; 0 Other;

XX

SQ Query Match 100.0%; Score 30; DB 4; Length 1584;

Best Local Similarity 100.0%; Pred. NO. 0.014;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCGGAGACCAAGGATGACCACAATGAG 30

DB 441 TGCGGAGACCAAGGATGACCACAATGAG 470

RESULT 2

AAI99682\_00/c

WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

| Fragment Name | Begin   | End     |
|---------------|---------|---------|
| AAI99682_00   | 1       | 110000  |
| AAI99682_01   | 100001  | 210000  |
| AAI99682_02   | 200001  | 310000  |
| AAI99682_03   | 300001  | 410000  |
| AAI99682_04   | 400001  | 510000  |
| AAI99682_05   | 500001  | 610000  |
| AAI99682_06   | 600001  | 710000  |
| AAI99682_07   | 700001  | 810000  |
| AAI99682_08   | 800001  | 910000  |
| AAI99682_09   | 900001  | 1010000 |
| AAI99682_10   | 1000001 | 1110000 |
| AAI99682_11   | 1100001 | 1210000 |
| AAI99682_12   | 1200001 | 1310000 |
| AAI99682_13   | 1300001 | 1410000 |
| AAI99682_14   | 1400001 | 1510000 |
| AAI99682_15   | 1500001 | 1610000 |
| AAI99682_16   | 1600001 | 1710000 |

|       |      |      |        |   |             |                 |           |     |
|-------|------|------|--------|---|-------------|-----------------|-----------|-----|
| 243   | 17.8 | 59.3 | 1886   | 4 | AAS31016    | Aas31016        | Human     | dis |
| 244   | 17.8 | 59.3 | 1935   | 5 | AAS84608    | Aas84608        | Drosophil |     |
| C 245 | 17.8 | 59.3 | 2249   | 4 | ABL13453    | ABL13453        | Drosophil |     |
| C 246 | 17.8 | 59.3 | 2303   | 6 | AAD36082    | Aad36082        | Mouse     | ATP |
| 247   | 17.8 | 59.3 | 2317   | 2 | AAQ56106    | DNA             | encod     |     |
| 248   | 17.8 | 59.3 | 2673   | 1 | AAN82199    | ENVN            | seq       |     |
| 249   | 17.8 | 59.3 | 2673   | 6 | AAX522308   | DNA             | encod     |     |
| 250   | 17.8 | 59.3 | 2673   | 6 | ABX52305    | CDNA            | encod     |     |
| 251   | 17.8 | 59.3 | 2673   | 8 | ADA07971    | DNA             | encod     |     |
| 252   | 17.8 | 59.3 | 3125   | 7 | ABX63016    | Human           | cdN       |     |
| C 253 | 17.8 | 59.3 | 3776   | 4 | ABL07232    | Drosophil       |           |     |
| C 254 | 17.8 | 59.3 | 4359   | 4 | ABU10138    | Drosophil       |           |     |
| 255   | 17.8 | 59.3 | 4382   | 4 | ABL13452    | Drosophil       |           |     |
| 256   | 17.8 | 59.3 | 5146   | 9 | ADD48890    | Human           | gen       |     |
| 257   | 17.8 | 59.3 | 5581   | 2 | AAT45451    | Human           | LST       |     |
| 258   | 17.8 | 59.3 | 5581   | 2 | AAT48737    | Human           | leu       |     |
| C 259 | 17.8 | 59.3 | 5969   | 4 | AAT32543    | Brevibact       |           |     |
| C 260 | 17.8 | 59.3 | 6911   | 2 | AAT34177    | Corynefor       |           |     |
| 261   | 17.8 | 59.3 | 9630   | 3 | AAB289628   | HIV-2           | ROD       |     |
| 262   | 17.8 | 59.3 | 9643   | 1 | AAN80859    | Sequence        |           |     |
| 263   | 17.8 | 59.3 | 9663   | 3 | AAC81938    | Packaging       |           |     |
| 264   | 17.8 | 59.3 | 9671   | 8 | ADA08001    | Human           | imm       |     |
| 265   | 17.8 | 59.3 | 9672   | 2 | AAQ20616    | ROD             | HIV-2     |     |
| 266   | 17.8 | 59.3 | 9726   | 3 | AAC81937    | Packaging       |           |     |
| 267   | 17.8 | 59.3 | 15424  | 4 | ABL27214    | Drosophil       |           |     |
| 268   | 17.8 | 59.3 | 38771  | 4 | AAK65890    | Human           | imm       |     |
| 269   | 17.8 | 59.3 | 48300  | 5 | AAF61281    | N. magada       |           |     |
| C 270 | 17.8 | 59.3 | 50000  | 3 | AAA96365    | Polymorph       |           |     |
| C 271 | 17.8 | 59.3 | 71428  | 7 | ACF65373    | Photorhab       |           |     |
| C 272 | 17.8 | 59.3 | 81800  | 6 | ABK94756    | Human           | cdN       |     |
| 273   | 17.8 | 59.3 | 110000 | 7 | ACF67367_05 | Continuation (6 | Of        |     |
| 274   | 17.8 | 59.3 | 113604 | 8 | ADA42607    | Human           | LQT       |     |
| 275   | 17.8 | 59.3 | 113604 | 8 | ADA42606    | Human           | LQT       |     |
| C 276 | 17.8 | 59.3 | 133894 | 2 | AAT13635    | ACNFV           | gen       |     |
| 277   | 17.8 | 59.3 | 349980 | 5 | AAH68533    | C               | Glutam    |     |
| C 278 | 17.6 | 58.7 | 65     | 9 | ADD68665    | DNA             | ampli     |     |
| C 279 | 17.6 | 58.7 | 177    | 4 | ABL24709    | Drosophil       |           |     |
| C 280 | 17.6 | 58.7 | 189    | 9 | ADD58670    | DNA             | ampli     |     |
| C 281 | 17.6 | 58.7 | 195    | 9 | ADD68667    | DNA             | ampli     |     |
| C 282 | 17.6 | 58.7 | 359    | 6 | ABG69588    | Novel           | mur       |     |
| 283   | 17.6 | 58.7 | 400    | 3 | AAZ94094    | Haematopo       |           |     |
| 284   | 17.6 | 58.7 | 410    | 4 | AAS38745    | Novel           | hum       |     |
| 285   | 17.6 | 58.7 | 514    | 5 | AAH64877    | Human           | sec       |     |
| C 286 | 17.6 | 58.7 | 576    | 3 | AAF08372    | Fusarium        |           |     |
| C 287 | 17.6 | 58.7 | 577    | 9 | ADC75098    | N               | bentham   |     |
| C 288 | 17.6 | 58.7 | 602    | 9 | ADD68692    | DNA             | ampli     |     |
| C 289 | 17.6 | 58.7 | 631    | 4 | AAS31316    | Human           | cdN       |     |
| C 290 | 17.6 | 58.7 | 631    | 4 | ABK43756    | DNA             | encod     |     |
| C 291 | 17.6 | 58.7 | 631    | 6 | ABQ66640    | Human           | pol       |     |
| C 292 | 17.6 | 58.7 | 631    | 9 | ADC10662    | Human           | cdN       |     |
| C 293 | 17.6 | 58.7 | 647    | 9 | ADD68691    | DNA             | ampli     |     |
| C 294 | 17.6 | 58.7 | 707    | 9 | ADD68690    | DNA             | ampli     |     |
| C 295 | 17.6 | 58.7 | 708    | 4 | ABL27243    | Drosophil       |           |     |
| C 296 | 17.6 | 58.7 | 712    | 4 | AAH07043    | Human           | cdN       |     |
| C 297 | 17.6 | 58.7 | 733    | 9 | ADD16812    | DNA             | (Seq)     |     |
| C 298 | 17.6 | 58.7 | 765    | 2 |             |                 |           |     |

## ALIGNMENTS

RESULT 1  
AAH51956  
ID AAH51956 standard; DNA; 1584 BP.  
XX  
XX AC AAH51956;  
XX  
XX DT 04-SEP-2001 (first entry)  
XX  
XX DE Mycobacterium tuberculosis potential drug target gene SEQ ID 10.  
XX

WP AAI99682\_17 1700001 1810000  
 WP AAI99682\_18 1800001 1910000  
 WP AAI99682\_19 1900001 2010000  
 WP AAI99682\_20 2000001 2110000  
 WP AAI99682\_21 2100001 2210000  
 WP AAI99682\_22 2200001 2310000  
 WP AAI99682\_23 2300001 2410000  
 WP AAI99682\_24 2400001 2510000  
 WP AAI99682\_25 2500001 2610000  
 WP AAI99682\_26 2600001 2710000  
 WP AAI99682\_27 2700001 2810000  
 WP AAI99682\_28 2800001 2910000  
 WP AAI99682\_29 2900001 3010000  
 WP AAI99682\_30 3000001 3110000  
 WP AAI99682\_31 3100001 3210000  
 WP AAI99682\_32 3200001 3310000  
 WP AAI99682\_33 3300001 3410000  
 WP AAI99682\_34 3400001 3510000  
 WP AAI99682\_35 3500001 3610000  
 WP AAI99682\_36 3600001 3710000  
 WP AAI99682\_37 3700001 3810000  
 WP AAI99682\_38 3800001 3910000  
 WP AAI99682\_39 3900001 4010000  
 WP AAI99682\_40 4000001 4110000  
 WP AAI99682\_41 4100001 4210000  
 WP AAI99682\_42 4200001 4310000  
 WP AAI99682\_43 4300001 4410000  
 WP AAI99682\_44 4400001 4411529  
 ID AAI99682 standard; DNA; 4411529 BP.  
 XX  
 AC AAI99682;  
 XX  
 AC  
 DT 15-JAN-2002 (first entry)  
 XX  
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
 XX  
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX  
 OS Mycobacterium tuberculosis.  
 XX  
 DN US6294328-B1.  
 XX  
 XX  
 PD 25-SEP-2001.  
 XX  
 PF 24-JUN-1998; 98US-00103840.  
 XX  
 PR 24-JUN-1998; 98US-00103840.  
 XX  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
 DR WPI; 2001-647261/74.  
 XX  
 DR  
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC 1551  
 PT and H37Rv differ.  
 XX  
 PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
 XX  
 CC The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen.  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly

CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1  
 XX  
 SQ Sequence 4411529 BP; 758565A; 1449983C; 1444602G; 758379T; 0U; 0Other;  
 Query Match 100.0%; Score 30; DB 4; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 0.026;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGGCGCAGAACCGAGTAGCACCACATGAG 30  
 DB 25004 TGGCGCAGAACCGAGTAGCACCACATGAG 24975

RESULT 3  
 AAI99683\_00/c  
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683  
 WP Fragment Name Begin End  
 WP AAI99683\_00 1 110000  
 WP AAI99683\_01 100001 210000  
 WP AAI99683\_02 200001 310000  
 WP AAI99683\_03 300001 410000  
 WP AAI99683\_04 400001 510000  
 WP AAI99683\_05 500001 610000  
 WP AAI99683\_06 600001 710000  
 WP AAI99683\_07 700001 810000  
 WP AAI99683\_08 800001 910000  
 WP AAI99683\_09 900001 1010000  
 WP AAI99683\_10 1000001 1110000  
 WP AAI99683\_11 1100001 1210000  
 WP AAI99683\_12 1200001 1310000  
 WP AAI99683\_13 1300001 1410000  
 WP AAI99683\_14 1400001 1510000  
 WP AAI99683\_15 1500001 1610000  
 WP AAI99683\_16 1600001 1710000  
 WP AAI99683\_17 1700001 1810000  
 WP AAI99683\_18 1800001 1910000  
 WP AAI99683\_19 1900001 2010000  
 WP AAI99683\_20 2000001 2110000  
 WP AAI99683\_21 2100001 2210000  
 WP AAI99683\_22 2200001 2310000  
 WP AAI99683\_23 2300001 2410000  
 WP AAI99683\_24 2400001 2510000  
 WP AAI99683\_25 2500001 2610000  
 WP AAI99683\_26 2600001 2710000  
 WP AAI99683\_27 2700001 2810000  
 WP AAI99683\_28 2800001 2910000  
 WP AAI99683\_29 2900001 3010000  
 WP AAI99683\_30 3000001 3110000  
 WP AAI99683\_31 3100001 3210000  
 WP AAI99683\_32 3200001 3310000  
 WP AAI99683\_33 3300001 3410000  
 WP AAI99683\_34 3400001 3510000  
 WP AAI99683\_35 3500001 3610000  
 WP AAI99683\_36 3600001 3710000  
 WP AAI99683\_37 3700001 3810000  
 WP AAI99683\_38 3800001 3910000  
 WP AAI99683\_39 3900001 4010000  
 WP AAI99683\_40 4000001 4110000  
 WP AAI99683\_41 4100001 4210000  
 WP AAI99683\_42 4200001 4310000  
 WP AAI99683\_43 4300001 4403765  
 ID AAI99683 standard; DNA; 4403765 BP.  
 XX  
 AC AAI99683;  
 XX  
 DT 15-JAN-2002 (first entry)  
 XX  
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.  
 XX  
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX  
 OS Mycobacterium tuberculosis.  
 OS

XX PN US6294328-B1.  
XX PD 25-SEP-2001.  
XX PF 24-JUN-1998; 98US-00103840.  
XX PR 24-JUN-1998; 98US-00103840.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX WPI; 2001-647261/74.  
XX  
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
XX determining the nucleotide sequence of the strain at positions in the  
XX genome corresponding to positions where M. tuberculosis strains CDC 1551  
XX and H37Rv differ.  
XX  
XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
XX  
XX The invention relates to evaluating strain variation within and between  
XX different populations of the tuberculosis bacterial pathogen,  
XX Mycobacterium tuberculosis or related Mycobacterium by determining the  
XX nucleotide sequence of the first strain at positions in the complete  
XX sequence of the genome that correspond to positions that differ in the  
XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
XX H37Rv (AAI99682). The method is useful for evaluating strain variation of  
XX M. tuberculosis and has valuable application in the fields of  
XX tuberculosis genetics, epidemiology, patient treatment and epidemic  
XX monitoring. Note: The sequence data for this patent did not form part of  
XX the printed specification, but was obtained in electronic format directly  
XX from USPTO at seqdata.uspto.gov/sequence.html?docID=6294328B1  
XX  
XX Sequence 4403765 BP; 757105A; 1447799C; 1441301G; 757371T; 0U; 1890Other;  
XX  
XX Query Match 100.0%; Score 30; DB 4; Length 110000;  
XX Best Local Similarity 100.0%; Pred. No. 0.026;  
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 TGGCGCAGAACGAGGTAGCACCAATGAG 30  
XX DB 24986 TGGCGCAGAACGAGGTAGCACCAATGAG 24957  
XX  
XX  
XX RESULT 4  
XX AAQ87723/c  
XX ID AAQ87723 standard; cDNA; 1476 BP.  
XX AC AAQ87723;  
XX DT 25-MAR-2003 (revised)  
XX DT 15-NOV-1995 (first entry)  
XX  
XX Human auxillary cytochrome P450 species 2B6 coding region.  
XX  
XX Human cytochrome P450; amplification; PCR; primer; expression vector;  
XX Yeast NADPH-P450 reductase; safety; fusion protein; metabolite;  
XX carcinogen; mutagen; liver metabolism; ds.  
XX  
XX Homo sapiens.  
XX  
XX EP644267-A2.  
XX PN 22-MAR-1995.  
XX PD 20-JUL-1994; 94EP-00111298.  
XX PF 20-JUL-1993; 93JP-00201120.  
XX PR 21-JUL-1993; 93JP-00180246.  
XX PR 30-JUL-1993; 93JP-00208279.  
XX

PA (SUMO) SUMITOMO CHEM CO LTD.  
XX Hayashi K, Sakaki T, Yabuseki Y, Komai K, Kareko H, Nakatsuka I;  
XX WPI; 1995-116991/16.  
XX DR P-PSDB; AAR72369.  
XX  
XX Evaluation of safety of a chemical cpd. - using recombinant yeast  
XX expressing human cytochrome P450 and a yeast NADPH-P450 reductase.  
XX  
XX Example; Page 57-59; 124pp; English.  
XX  
XX The nucleotide sequence of the cDNA coding region for the human auxillary  
XX cytochrome P450 species 2B6. The gene encodes a protein of 491 amino  
XX acids. The cDNA was amplified by PCR using the primers AAQ87755-6. The  
XX product was cloned into the yeast expression vectors pAAH5N or pAHR5 to  
XX produce the vectors p2B6 for the expression of the cytochrome P450 alone  
XX or p2B6R for co-expression with the yeast NADPH-P450 reductase. The  
XX vectors are used in a method for evaluating the safety of a chemical  
XX compound by reacting the chemical compound with recombinantly produced  
XX human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),  
XX 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxillary species and variants  
XX or in cell extracts, and analysing the resulting metabolite to assess the  
XX safety of the chemical compound. The method is useful for determining  
XX whether the chemical compound, or its metabolite, will be converted into  
XX a carcinogenic or mutagenic form through metabolism in the liver.  
XX (Updated on 25-MAR-2003 to correct PN field.)  
XX  
XX Sequence 1476 BP; 351 A; 453 C; 338 G; 334 T; 0 U; 0 Other;  
XX  
XX Query Match 68.7%; Score 20.6; DB 2; Length 1476;  
XX Best Local Similarity 85.2%; Pred. No. 1e+02;  
XX Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
XX  
XX QY 1 TGGCGCAGAACGAGGTAGCACCAAT 27  
XX DB 68 TGGCGCTGAACGAGGTAGCACCAAT 42  
XX  
XX  
XX RESULT 5  
XX AAT28389/c  
XX ID AAT28389 standard; DNA; 1476 BP.  
XX AC AAT28389;  
XX DT 11-OCT-1996 (first entry)  
XX DE Human cytochrome P450 molecular species 2B6 gene.  
XX  
XX Human cytochrome P450; amplified; PCR; polymerase chain reaction; primer;  
XX liver; yeast; expression vector; NADPH-P450 reductase; ADH gene promoter;  
XX evaluation; safety; fusion protein; metabolite; detoxification;  
XX carcinogenic; ds.  
XX  
XX Homo sapiens.  
XX  
XX JP08056695-A.  
XX PN 05-MAR-1996.  
XX PD 15-JUL-1994; 94JP-00164184.  
XX PF 20-JUL-1993; 93JP-00201120.  
XX PR 30-JUL-1993; 93JP-00208279.  
XX PR 17-JUN-1994; 94JP-00136053.  
XX  
XX (SUMO) SUMITOMO CHEM CO LTD.  
XX WPI; 1996-182311/19.  
XX DR P-PSDB; AAR93176.  
XX  
XX Novel method for the evaluation of the safety of a cpd. - using a human  
XX

PT cytochrome P450 and yeast NADPH reductase to determine whether the  
 PT analyte cpd. is detoxified or metabolised to a carcinogen.

PS Example 1; Page 37-39; 74pp; Japanese.

XX This is the nucleotide sequence of the human cytochrome P450 molecular  
 CC species 2B6 gene which encodes a protein of 491 amino acids. The 1.5 kb  
 CC fragment was amplified from a human liver derived cDNA library using  
 CC primers AAT26945-6. The prod. was cloned into the yeast expression vector  
 CC pAAH5N to generate plasmid p2B6 for prodn. of the cytochrome only or into  
 CC the vector pAHR to generate the plasmid p2B6R for co-prodn. with the  
 CC yeast NADPH-P450 reductase. The sequence is placed under control of the  
 CC yeast ADH gene promoter and terminator. The vectors are used in a method  
 CC for evaluating the safety of a cpd. by reacting the test cpd. with  
 CC recombinantly produced human cytochrome P450 mol. species 1A2 (AAT28380),  
 CC 2C9 (28381), 2E1 (AAT28382), 3A4 (AAT28383) or their variants (AAT28384-  
 CC 98) together with yeast NADPH-P450 reductase (either as a fused protein  
 CC or as a cell extract) and analysing the resultant metabolite. The cpd. is  
 CC considered "safe" if it is detoxified or not rendered carcinogenic or  
 CC "unsafe" if it is not detoxified or is metabolised to a carcinogenic cpd

XX Sequence 1476 BP; 351 A; 453 C; 338 G; 334 T; 0 U; 0 Other;

Query Match 68.7%; Score 20.6; DB 2; Length 1476;  
 Best Local Similarity 85.2%; Pred. No. 1e+02; Mismatches 0; Gaps 0;  
 Matches 23; Conservative 0; Indels 4;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
 DB 68 TGGCGCTGAACCCAGGAGTAGCAGAGT 42

## RESULT 6

AAT17417/C  
 ID AAT17417 standard; cDNA; 1476 BP.

AC AAT17417;

DT 01-AUG-1996 (first entry)

XX Human derived cytochrome P4502B6 cDNA.

DE Human derived cytochrome; P4502B6; commercial cDNA library; yeast;  
 KW transfection; recombinant production; expression vector; mammal;  
 KW immunisation; sensitisation; antibody; determination; detection;  
 KW non-cross reactive; ds.

XX Homo sapiens.

FT Key Location/Qualifiers  
 FT CDS 1..1476  
 FT /\*tag= a

FT JP08027193-A.

PN 30-JAN-1996.

XX 13-JUL-1994; 94JP-00161548.

XX 13-JUL-1994; 94JP-00161548.

XX (SUMO ) SUMITOMO CHEM CO LTD.

XX WPI: 1996-136334/14.

DR P-PSDB; AAR81468.

XX Antibody recognising human derived cytochrome P4502B6 - allows specific  
 PT detection of cytochrome P450 species in humans.

PS Example 1; Page 10-12; 13pp; Japanese.

XX The present sequence encodes the human derived cytochrome (HDC) P4502B6,  
 CC which was obtd. from a commercial cDNA library. Yeast were transfected

CC with an expression vector contg. the HDC cDNA, cultured and then  
 CC disrupted to give a microsomal fraction. The HDC was purified from the  
 CC fraction, and used to immunise and sensitise a mammal. Blood was drawn  
 CC from the mammal, and an anti-HDC antibody isolated. The antibody obtd.  
 CC recognises HDC P4502B6, partic. at a serum dilution rate of 1:10000, and  
 CC is substantially without cross reaction to other HDC P450 spp

XX Sequence 1476 BP; 351 A; 453 C; 338 G; 334 T; 0 U; 0 Other;

Query Match 68.7%; Score 20.6; DB 2; Length 1476;  
 Best Local Similarity 85.2%; Pred. No. 1e+02; Mismatches 0; Gaps 0;  
 Matches 23; Conservative 0; Indels 4;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
 DB 68 TGGCGCTGAACCCAGGAGTAGCAGAGT 42

## RESULT 7

AAS11662/c

ID AAS11662 standard; DNA; 2915 BP.

XX AC AAS11662;

XX DT 24-OCT-2001 (first entry)

XX DE Human CYP2B6 DNA 5' flanking sequence.

XX CYP2B6; cytostatic; gene therapy; genotyping; cancer; metabolism; ds;  
 KW human; cancer susceptibility; environmental carcinogen;  
 KW 5' flanking sequence.

XX Homo sapiens.

FT Key Location/Qualifiers  
 FT primer\_bind complement(1..18)  
 FT /\*tag= a  
 FT /bound moiety= "PCR primer CYP2B6-P2F"  
 FT primer\_bind complement(565..584)  
 FT /\*tag= b  
 FT /bound moiety= "Sequencing primer seqCYP2B6-P1F"  
 FT variation replace(688,G)  
 FT /\*tag= c  
 FT variation replace(1010,C)  
 FT /\*tag= d  
 FT primer\_bind complement(1212..1232)  
 FT /\*tag= e  
 FT /bound moiety= "Sequencing primer seqCYP2B6-P2F"  
 FT variation replace(1280,G)  
 FT /\*tag= f  
 FT variation replace(1715,C)  
 FT /\*tag= g  
 FT primer\_bind complement(1865..1883)  
 FT /\*tag= h  
 FT /bound moiety= "Sequencing primer seqCYP2B6-P3F"  
 FT primer\_bind 1944..1964  
 FT /\*tag= i  
 FT /bound moiety= "Sequencing primer seqCYP2B6-P4R"  
 FT primer\_bind complement(2193..2210)  
 FT /\*tag= j  
 FT /bound moiety= "PCR primer CYP2B6-P1F"  
 FT primer\_bind 2193..2210  
 FT /\*tag= k  
 FT /bound moiety= "PCR primer CYP2B6-P1R"  
 FT variation replace(2383,C)  
 FT /\*tag= l  
 FT exon 2455..2550  
 FT /\*tag= m  
 FT /number= 1  
 FT 2507..2526  
 FT primer\_bind /\*tag= n  
 FT /bound moiety= "Sequencing primer seqCYP2B6-P5R"  
 FT primer\_bind 2898..2915

```

FT      /*tag= o
FT      /bound_moisty= "PCR primer CYP2B6-PrR"
XX      WO200159152-A2.
XX      PD      16-AUG-2001.
XX      XX
XX      PF      09-FEB-2001; 2001WO-EP001456.
XX      XX
XX      PR      09-FEB-2000; 2000EP-00102701.
XX      XX
XX      PA      (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX      XX
XX      PI      Zanger UM, Lang T;
XX      XX
XX      DR      WPI; 2001-502719/55.
XX      XX
XX      PT      New polynucleotide(s) of the polymorphic human CYP2B6 gene for the
XX      PT      detection and treatment of disorders i.e. cancer.
XX      PS      Example 2; Fig 2; 83pp; English.
XX      CC      The sequence represents the 5' flanking sequence of human CYP2B6 DNA.
XX      CC      CYP2B6 polynucleotides and oligonucleotides are used for specific
XX      CC      detection and genotyping of CYP2B6 alleles in humans, determination of
XX      CC      which is useful for the optimisation of therapies utilising CYP2B6
XX      CC      substrates. Oligonucleotide sequences are useful in detection of the
XX      CC      individual predisposition to several common cancers caused by
XX      CC      environmental carcinogens, and diseases treated with drugs that are
XX      CC      targets of the CYP2B6 gene product, whose metabolism is therefore
XX      CC      dependent on CYP2B6. Cancer or susceptibility to cancer can be diagnosed
XX      CC      by detecting the presence of a molecular variant of CYP2B6. From variants
XX      CC      of the alleles, modulators of the activity can be developed for use in
XX      CC      treatment and prevention of CYP2B6-related disorders
XX      SQ      Sequence 2915 BP; 812 A; 714 C; 645 G; 744 T; 0 U; 0 Other;

Query Match      68.7%; Score 20.6; DB 4; Length 2915;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGGCGCAGACACGAGGTAGCACCAAT 27
Db      2532 TGGCGCTGACACGAGGTAGCAAGAGT 2506

RESULT 8
AAZ19787/c
ID      AAZ19787 standard; cDNA; 3045 BP.
XX      AC      AAZ19787;
XX      DT      06-DEC-1999 (first entry)
XX      DE      Human cytochrome P450 2B6 cDNA.
XX      KW      Cytochrome; targeting; localisation; cancer; tumour; produg; reduction;
XX      KW      ds.
XX      OS      Homo sapiens.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      7..1482
XX      FT      /*tag= a
XX      FT      /product= "Human cytochrome P450 2B6 CYP2B6"
XX      FT      intron 1266..1446
XX      FT      /*tag= b
XX      PN      WO9945127-A2.
XX      XX
XX      PD      10-SEP-1999.
XX      XX
XX      PF      05-MAR-1999; 99WO-GB000674.

XX      06-MAR-1998; 98GB-00004841.
XX      PR      19-AUG-1998; 98GB-00018103.
XX      PR      29-JAN-1999; 99GB-00002081.
XX      XX
XX      FA      (OXFO-) OXFORD BIOMEDICA UK LTD.
XX      XX
XX      PI      Stratford IJ, Patterson AV, Kingsman SM, Kan O, Griffiths L;
XX      PI      Mitrophanous K;
XX      XX
XX      DR      WPI; 1999-551046/46.
XX      DR      P-PSDB; AAY42295.
XX      XX
XX      PT      New produg activating agent targeted to selected cells or tissues,
XX      PT      particularly hypoxic cells, for treating e.g. tumors.
XX      XX
XX      PS      Example 12; Fig 4; 187pp; English.
XX      CC      This sequence represents human cytochrome P450 2B6 (CYP2B6) cDNA. CYP2B6,
XX      CC      and/or cytochrome P450 reductase (P450R) can be used to activate produgs
XX      CC      to their active form via reduction. Administration of a produg is useful
XX      CC      where the active drug may be metabolised before it reaches its site of
XX      CC      action or where the active drug is cytotoxic, e.g., anticancer drugs.
XX      CC      Cytochrome P450 isoforms such as CYP2B6, or vectors that express them,
XX      CC      can be used to treat tumours, inflammation, atherosclerosis and muscular
XX      CC      dystrophy, and may also be used to treat many other conditions, e.g.,
XX      CC      cerebral malaria, rheumatoid arthritis, or conditions associated with
XX      CC      hypoxia, ischaemia or hypoglycemia, or to deliver antibiotics, antiviral
XX      CC      agents, analgesics, anaesthetics, anti-inflammatory, antineoplastic
XX      CC      agents and diagnostic agents
XX      SQ      Sequence 3045 BP; 706 A; 864 C; 661 G; 814 T; 0 U; 0 Other;

Query Match      68.7%; Score 20.6; DB 2; Length 3045;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGGCGCAGACACGAGGTAGCACCAAT 27
Db      74 TGGCGCTGACACGAGGTAGCAAGAGT 48

RESULT 9
AAZ07812/c
ID      AAZ07812 standard; DNA; 3045 BP.
XX      AC      AAZ07812;
XX      DT      23-NOV-1999 (first entry)
XX      DE      Human cytochrome P450 2B6 nucleotide sequence.
XX      KW      Produg; localization domain; tumor-selective antibody; cytochrome P450;
XX      KW      produg activating domain; modified hematopoietic stem cell; MHSC; tumor;
XX      KW      inflammation; atherosclerosis; muscular dystrophy; cerebral malaria;
XX      KW      rheumatoid arthritis; hypoxia; ischemia; hypoglycemia; human; ss.
XX      OS      Homo sapiens.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      7..1482
XX      FT      /*tag= a
XX      FT      /transl_except= (pos:538..540, aa: Ile)
XX      FT      /transl_except= (pos:925..927, aa: Leu)
XX      FT      misc_feature 1277..1446
XX      FT      /*tag= b
XX      FT      /note= "the amino acid residues encoded by this region is
XX      FT      not indicated in the corresponding protein"
XX      PN      WO9945126-A2.
XX      XX
XX      PD      10-SEP-1999.
XX      XX

```

PF 05-MAR-1999; 99WO-GB000672.  
XX  
PR 06-MAR-1998; 98GB-00004841.  
PR 19-AUG-1998; 98GB-00018103.  
XX 29-JAN-1999; 99GB-00002081.  
XX  
PA (OXFO-) OXFORD BIOMEDICA UK LTD.  
XX  
XX Stratford IJ, Patterson AV, Kingsman SM, Kan O, Griffiths L;  
PI Mitrophanous K;  
PI  
XX WPI: 1999-540852/45.  
XX P-PSDB; AA27406.  
XX  
PT New prodrug activating agent targeted to selected cells or tissues,  
PT particularly hypoxic cells, for treating e.g. tumors or inflammation.  
XX  
PS Example; Fig 4; 149pp; English.  
XX  
XX The invention provides a new prodrug activating agent that comprises: (i)  
CC a localization domain (LD; other than a tumor-selective antibody) and a  
CC prodrug activating domain (PAD); (ii) at least one nucleic acid encoding  
CC a cytochrome P450 and under control of at least one constitutive or  
CC inducible expression control sequence or (iii) a modified hematopoietic  
CC stem cell (WHSC) containing at least one nucleic acid encoding a PAD and  
CC under control of elements as in (ii). The prodrug activating agent or  
CC vectors that express them, are specifically used to treat tumors, or  
CC inflammation, atherosclerosis and muscular dystrophy, but may also be  
CC used to treat many other conditions, e.g. cerebral malaria, rheumatoid  
CC arthritis, or conditions associated with hypoxia, hypoglycemia or  
CC ischemia, or to deliver antibiotics, antiviral agents, analgesics,  
CC anesthetics, anti-inflammatory, antineoplastic agents and diagnostic  
CC agents. LD optimize activity of PAD, e.g. by delivering it to selected  
CC locations or by delivering it to neighboring cells (bystander effect),  
CC and allow a reduction in dose of prodrug, and thus of systemic side-  
CC effects. Nucleic acids encoding the agent may be expressed selectively in  
CC hypoxic cells. The present sequence represents the human cytochrome P450  
CC 2B6 nucleotide sequence  
XX  
SQ Sequence 3045 BP; 706 A; 864 C; 661 G; 814 T; 0 U; 0 Other;  
  
Query Match 68.7%; Score 20.6; DB 2; Length 3045;  
Best Local Similarity 85.2%; Pred. NO. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 TGCGCGAGAACGAGGAGTAGCCCAAT 27  
Db 74 TGCGCGTGACCGAGGAGTAGCCAGAGT 48  
  
RESULT 10  
ID ABL64908/c  
XX ABL64908 standard; DNA; 3045 BP.  
XX  
AC ABL64908;  
XX  
XX 15-MAY-2002 (first entry)  
DT  
XX  
XX Lung cancer related gene sequence SEQ ID NO:3245.  
DE  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200194629-A2.  
PN  
XX  
PD 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US010838.  
PF  
XX

PR 05-JUN-2000; 2000US-0209473P.  
PR 05-JUN-2000; 2000US-0209531P.  
PR 18-SEP-2000; 2000US-0233133P.  
PR 18-SEP-2000; 2000US-0233617P.  
PR 20-SEP-2000; 2000US-0234009P.  
PR 20-SEP-2000; 2000US-0234034P.  
PR 20-SEP-2000; 2000US-0234052P.  
PR 22-SEP-2000; 2000US-0234509P.  
PR 22-SEP-2000; 2000US-0234567P.  
PR 25-SEP-2000; 2000US-0234923P.  
PR 25-SEP-2000; 2000US-0234924P.  
PR 25-SEP-2000; 2000US-0235077P.  
PR 25-SEP-2000; 2000US-0235082P.  
PR 25-SEP-2000; 2000US-0235134P.  
PR 25-SEP-2000; 2000US-0235280P.  
PR 26-SEP-2000; 2000US-0235637P.  
PR 26-SEP-2000; 2000US-0235638P.  
PR 27-SEP-2000; 2000US-0235711P.  
PR 27-SEP-2000; 2000US-0235720P.  
PR 27-SEP-2000; 2000US-0235840P.  
PR 27-SEP-2000; 2000US-0235863P.  
PR 28-SEP-2000; 2000US-0236028P.  
PR 28-SEP-2000; 2000US-0236032P.  
PR 28-SEP-2000; 2000US-0236033P.  
PR 28-SEP-2000; 2000US-0236034P.  
PR 28-SEP-2000; 2000US-0236109P.  
PR 28-SEP-2000; 2000US-0236111P.  
PR 29-SEP-2000; 2000US-0236842P.  
PR 29-SEP-2000; 2000US-0236891P.  
PR 02-OCT-2000; 2000US-0237172P.  
PR 02-OCT-2000; 2000US-0237173P.  
PR 02-OCT-2000; 2000US-0237278P.  
PR 02-OCT-2000; 2000US-0237294P.  
PR 02-OCT-2000; 2000US-0237395P.  
PR 02-OCT-2000; 2000US-0237316P.  
PR 03-OCT-2000; 2000US-0237425P.  
PR 03-OCT-2000; 2000US-0237598P.  
PR 03-OCT-2000; 2000US-0237604P.  
PR 03-OCT-2000; 2000US-0237606P.  
PR 03-OCT-2000; 2000US-0237608P.  
PR 01-NOV-2000; 2000US-0244867P.  
PR 01-NOV-2000; 2000US-0245084P.  
XX  
FA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
DR  
XX  
PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
PT agent to be tested for anti-neoplastic activity, and determining a change  
PT in expression of a gene of a signature gene set.  
XX  
PS Claim 1; SEQ ID NO 3245; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an anti-  
CC neoplastic agent. The method involves exposing cells to a chemical agent  
CC to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening an  
CC anti-neoplastic agent, and can be used for producing a product which is  
CC the data collected with respect to the anti-neoplastic agent as a result  
CC of M1, and the data is sufficient to convey the chemical structure and/or  
CC properties of the agent. M1 can be used in the treatment of cancer such  
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
CC tumour



XX SQ Sequence 3045 BP; 706 A; 866 C; 658 G; 815 T; 0 U; 0 Other;  
Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
||||| ||||||| ||||||| ||||||| |||  
Db 74 TGGCGCTGACACGAGGAGTAGCAAGACT 48

RESULT 11  
ABT10910/c  
ID ABT10910 standard; cDNA; 3045 BP.  
XX AC ABT10910;  
XX DT  
XX DE 04-DEC-2002 (first entry)  
XX DE Human breast cancer associated coding sequence SEQ ID NO: 1044.  
XX KW Human; breast specific gene; breast cancer; differential expression;  
XX KW cytotstatic; gene therapy; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200259271-A2.  
XX PD 01-AUG-2002.  
XX PF 25-JAN-2002; 2002WO-US0021176.  
XX PR 25-JAN-2001; 2001US-0263757P.  
XX PR 25-APR-2001; 2001US-0286090P.  
XX PR 23-MAY-2001; 2001US-0292517P.  
XX PA (GENE-) GENE LOGIC INC.  
XX PI Orr MS, Nation M, Diggins JC, Zeng W;  
XX WPI; 2002-674803/72.  
XX PT Diagnosing breast cancer in a patient comprises detecting the level of  
XX PT gene expression in cell or tissue samples, where a differential gene  
XX PT expression is indicative of breast cancer.  
XX PS Claim 1; SEQ ID NO 1044; 260pp + Sequence Listing; English.  
XX CC The present invention relates to methods of diagnosing breast cancer in a  
XX CC patient, which comprise detecting the level of expression in a tissue  
XX CC sample of two or more genes selected from those shown in ABT09867-  
XX CC ABT1112, where a differential expression of the genes indicates breast  
XX CC cancer. The methods are useful in diagnosing, treating, detecting the  
XX CC progression, and in monitoring treatment of breast cancer in patients.  
XX CC The methods are also useful as a screening tool for agents that modulate  
XX CC the onset or progression of breast cancer. The breast cancer genes may be  
XX CC used as diagnostic markers for the prediction or identification of the  
XX CC malignant state of breast tissue, for confirming the type and progression  
XX CC of cancer, and for drug screening and assays. The present sequence is a  
XX CC coding sequence of the invention. Note: The sequence data for this patent  
XX CC did not form part of the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 3045 BP; 706 A; 866 C; 658 G; 815 T; 0 U; 0 Other;  
Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
||||| ||||||| ||||||| ||||||| |||  
Db 74 TGGCGCTGACACGAGGAGTAGCAAGACT 48

RESULT 13  
ADD29818/c  
ID ADD29818 standard; mRNA; 4887 BP.  
XX AC ADD29818;  
XX DT 15-JAN-2004 (first entry)

Db 74 TGGCGCTGACACGAGGAGTAGCAAGACT 48

RESULT 12  
ABN95821/c  
ID ABN95821 standard; DNA; 3045 BP.  
XX AC ABN95821;  
XX DT 13-AUG-2002 (first entry)  
XX DE Gene #2319 used to diagnose liver cancer.  
XX KW Gene: liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
XX KW metastatic liver tumour; cytostatic; expression profile; disease state;  
XX KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX OS Homo sapiens.  
XX PN WO200229103-A2.  
XX PD 11-APR-2002.  
XX PF 02-OCT-2001; 2001WO-US030589.  
XX PR 02-OCT-2000; 2000US-0237054P.  
XX PA (GENE-) GENE LOGIC INC.  
XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX WPI; 2002-426119/45.  
XX PT Diagnosing and detecting the progression of liver cancer, hepatocellular  
XX PT carcinoma or metastatic liver tumor in a patient, involves detecting the  
XX PT level of expression of two or more genes in a liver tissue sample.  
XX PS Claim 1; SEQ ID NO 2319; 298pp; English.  
XX CC The invention relates to a novel method for diagnosing and detecting the  
XX CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
XX CC tumour in a patient, and differentiating metastatic liver cancer from  
XX CC hepatocellular carcinoma in a patient, involving detecting the level of  
XX CC expression of two or more genes represented in ABN93503-ABN97455 in a  
XX CC tissue sample. The method of the invention has hepatotropic, and  
XX CC cytostatic activity. The method is useful for diagnosing and detecting  
XX CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
XX CC liver carcinoma in a patient. The method is useful for identifying  
XX CC expression profiles which serve as useful diagnostic markers as well as  
XX CC markers that can be used to monitor disease states, disease progression,  
XX CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
XX CC for this patent did not form part of the printed specification, but was  
XX CC obtained in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 3045 BP; 706 A; 866 C; 658 G; 815 T; 0 U; 0 Other;  
Query Match 68.7%; Score 20.6; DB 6; Length 3045;  
Best Local Similarity 85.2%; Pred. No. 1.1e+02;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
||||| ||||||| ||||||| ||||||| |||  
Db 74 TGGCGCTGACACGAGGAGTAGCAAGACT 48

RESULT 13  
ADD29818/c  
ID ADD29818 standard; mRNA; 4887 BP.  
XX AC ADD29818;  
XX DT 15-JAN-2004 (first entry)

```
XX DE Human tumour suppressor mRNA SEQ ID NO:304.
XX DE
XX ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.
XX KW
XX OS Homo sapiens.
XX PN WO2003058201-A2.
XX XX
XX PD 17-JUL-2003.
XX XX
XX PF 31-DEC-2002; 2002WO-US041825.
XX XX
XX PR 31-DEC-2001; 2001US-0345317P.
XX XX
XX PA (QUAR-) QUARK BIOTECH INC.
XX PA (CLEV-) CLEVELAND CLINIC FOUND.
XX PI Feinstein E, Gudkov AV;
XX PI
XX DR WPI; 2003-598393/56.
XX XX
XX PT Diagnosing cancer comprises determining the polypeptide or polynucleotide
XX PT levels e.g., hepatic lipase, in a sample from a subject, where a higher
XX PT level compared to that in a subject free of cancer is indicative of
XX PT cancer.
XX PS Disclosure; SEQ ID NO 304; 272pp; English.
XX XX
XX CC The invention relates to a novel method for diagnosing a cancer in a
XX CC subject. The method comprises determining, in a sample from the subject,
XX CC the level of at least one polypeptide, where a higher level of the
XX CC polypeptide compared to the level of the polypeptide in a subject free of
XX CC cancer is indicative of cancer. The polypeptide is selected from any of
XX CC the polypeptides encoded by the polynucleotides listed in the
XX CC specification and polypeptides which are at least 70% homologous to the
XX CC polypeptides. The method of the invention has cytostatic activity, and
XX CC may have a use in gene therapy. The method is useful in identifying
XX CC markers specific for one or several types of cancer, depending on the
XX CC tissue origin, which may be used in numerous diagnostic and prognostic
XX CC applications as well as cancer type-specific targets for therapeutic
XX CC intervention. The compounds that modulate the activity of a tumour
XX CC suppressor gene are useful in the treatment of cancer or as anti-cancer
XX CC drugs. The present sequence represents a polynucleotide of the invention.
XX XX
XX SQ Sequence 4887 BP; 1238 A; 1327 C; 1071 G; 1251 T; 0 U; 0 Other;
XX XX
Query Match 68.7%; Score 20.6; DB 9; Length 4887;
Best Local Similarity 85.2%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX QY 1 TGGCGCAGAACCGAGGAGTAGCAAT 27
XX DB 1910 TGGCGCTGAACCGAGGAGTAGCAAGT 1884
XX XX
RESULT 14
AAS11651/C
ID AAS11651 standard; DNA; 41 BP.
XX AC
XX AAS11651;
XX DT
XX 24-OCT-2001 (first entry)
XX DE Human CYP2B6 allele variant polynucleotide fragment #6.
XX XX
XX CYP2B6; cytostatic; gene therapy; genotyping; cancer; metabolism; ds;
XX KW human; cancer susceptibility; environmental carcinogen.
XX XX
XX OS Homo sapiens.
XX PN WO200159152-A2.
XX XX
XX PD 16-AUG-2001.
XX XX
XX PF 09-FEB-2001; 2001WO-EP001456.
XX XX
XX PR 09-FEB-2000; 2000EP-00102701.
XX XX
XX PA (EPID-) EPIDAUROS BIOTECHNOLOGIE AG.
XX XX
XX PI Zanger UM, Lang T;
XX PI
XX DR WPI; 2001-502719/55.
XX XX
XX PT New polynucleotide(s) of the polymorphic human CYP2B6 gene for the
XX PT detection and treatment of disorders i.e. cancer.
XX PS Claim 1; Page 80; 83pp; English.
XX XX
XX CC The sequence represents a polynucleotide variant sequence of the human
XX CC CYP2B6 gene. It is used for specific detection and genotyping of CYP2B6
XX CC alleles in humans, determination of which is useful for the optimisation
XX CC of therapies utilising CYP2B6 substrates. Oligonucleotide sequences are
XX CC useful in detection of the individual predisposition to several common
XX CC cancers caused by environmental carcinogens, and diseases treated with
XX CC drugs that are targets of the CYP2B6 gene product, whose metabolism is
XX CC therefore dependent on CYP2B6. Cancer or susceptibility to cancer can be
XX CC diagnosed by detecting the presence of a molecular variant of CYP2B6.
XX CC From variants of the alleles, modulators of the activity can be developed
XX CC for use in treatment and prevention of CYP2B6-related disorders
XX XX
XX SQ Sequence 41 BP; 8 A; 17 C; 6 G; 10 T; 0 U; 0 Other;
XX XX
Query Match 68.0%; Score 20.4; DB 4; Length 41;
Best Local Similarity 95.8%; Pred. No. 72;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX QY 1 TGGCGCAGAACCGAGGAGTAGCA 22
XX DB 25 TGGCGCTGAACCGAGGAGTAGCA 4
XX XX
RESULT 15
AAH69361
ID AAH69361 standard; cDNA; 590 BP.
XX AC
XX AAH69361;
XX XX
XX DT 19-SEP-2001 (first entry)
XX XX
XX DE Human cervical cancer marker nucleic acid 635.
XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX XX
XX OS Homo sapiens.
XX PN WO200142467-A2.
XX XX
XX PD 14-JUN-2001.
XX XX
XX PF 08-DEC-2000; 2000WO-US033312.
XX XX
XX PR 08-DEC-1999; 99US-0169681P.
XX PR 21-DEC-1999; 99US-0171350P.
XX PR 14-MAR-2000; 2000US-0189315P.
XX PR 12-MAY-2000; 2000US-0203791P.
XX PR 09-JUN-2000; 2000US-0210600P.
XX PR 21-JUL-2000; 2000US-0220114P.
XX XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PA
XX PI Schlegel R, Deeds J, Berger A, Zhao X;
XX XX
XX DR WPI; 2001-375006/39.
XX XX
```

PT New isolated nucleic acid for diagnosing and treating cervical cancer and  
PT for assessing and detecting compounds for treating the cancer.

XX PS Claim 1; Page 214; 1051pp; English.

XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy

XX SQ Sequence 590 BP; 227 A; 110 C; 98 G; 153 T; 0 U; 2 Other;

Query Match 65.3%; Score 19.6; DB 4; Length 590;

Best Local Similarity 84.6%; Pred. No. 2.3e+02;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CGCAGAACCCAGGAGTAGCACCACCAATGA 29

DB 409 CACAGAACCCAGGAGTAGCACCACCAATCA 434

RESULT 16

AAH72227

ID AAH72227 standard; cDNA; 649 BP.

XX AC AAH72227;

XX DT 19-SEP-2001 (first entry)

XX DE Human cervical cancer marker nucleic acid 3501.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US033312.

XX PR 08-DEC-1999; 99US-0169681P.

XX PR 21-DEC-1999; 99US-0171350P.

XX PR 14-MAR-2000; 2000US-0189315P.

XX PR 12-MAY-2000; 2000US-0203791P.

XX PR 09-JUN-2000; 2000US-0210600P.

XX PR 21-JUL-2000; 2000US-0220114P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX PT New isolated nucleic acid for diagnosing and treating cervical cancer and

XX PT for assessing and detecting compounds for treating the cancer.

XX PS Claim 1; Page 669; 1051pp; English.

XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy

XX SQ Sequence 649 BP; 253 A; 118 C; 108 G; 170 T; 0 U; 0 Other;

Query Match 65.3%; Score 19.6; DB 4; Length 649;

Best Local Similarity 84.6%; Pred. No. 2.3e+02;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CGCAGAACCCAGGAGTAGCACCACCAATGA 29

DB 407 CACAGAACCCAGGAGTAGCACCACCAATCA 432

RESULT 17

AAH71164

ID AAH71164 standard; cDNA; 778 BP.

XX AC AAH71164;

XX DT 19-SEP-2001 (first entry)

XX DE Human cervical cancer marker nucleic acid 2438.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US033312.

XX PR 08-DEC-1999; 99US-0169681P.

XX PR 21-DEC-1999; 99US-0171350P.

XX PR 14-MAR-2000; 2000US-0189315P.

XX PR 12-MAY-2000; 2000US-0203791P.

XX PR 09-JUN-2000; 2000US-0210600P.

XX PR 21-JUL-2000; 2000US-0220114P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX PT New isolated nucleic acid for diagnosing and treating cervical cancer and

XX PT for assessing and detecting compounds for treating the cancer.

XX PS Claim 1; Page 511-512; 1051pp; English.

XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy

XX SQ Sequence 778 BP; 280 A; 153 C; 135 G; 207 T; 0 U; 3 Other;

Query Match 65.3%; Score 19.6; DB 4; Length 778;

Best Local Similarity 84.6%; Pred. No. 2.4e+02;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 CGCAGAACCCAGGAGTAGCACCACCAATGA 29

DB 420 CACAGAACCCAGGAGTAGCACCACCAATCA 445

RESULT 18

AAH72859/c

ID AAH72859 standard; cDNA; 2420 BP.

XX AC AAH72859;

XX 19-SEP-2001 (first entry)  
XX Human cervical cancer marker nucleic acid 4133.  
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX Homo sapiens.  
XX WO200142467-A2.  
XX 14-JUN-2001.  
XX 08-DEC-2000; 2000WO-US033312.  
XX 08-DEC-1999; 99US-0169681P.  
XX 21-DEC-1999; 99US-0171350P.  
XX 14-MAR-2000; 2000US-0189315P.  
XX 13-MAY-2000; 2000US-0203791P.  
XX 09-JUN-2000; 2000US-0210600P.  
XX 21-JUL-2000; 2000US-0220114P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Schlegel R, Deeds J, Berger A, Zhao X;  
XX WPI; 2001-375006/39.  
XX New isolated nucleic acid for diagnosing and treating cervical cancer and  
XX PT for assessing and detecting compounds for treating the cancer.  
XX PS Claim 1; Page 894; 1051pp; English.  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
XX CC polypeptides are useful to assess if a patient is afflicted with  
XX CC cervical cancer or has a pre-malignant condition; to monitor the  
XX CC progression of cervical cancer or a premalignant condition in a patient;  
XX CC and to select and/or assess the efficacy of a compound or therapy for  
XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
XX CC useful for gene therapy  
XX SQ Sequence 2420 BP; 641 A; 458 C; 473 G; 838 T; 0 U; 10 Other;  
Query Match 65.3%; Score 19.6; DB 4; Length 2420;  
Best Local Similarity 84.6%; Pred. No. 2.8e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 4 CGCAGAACCCAGGAGTACCCCATGA 29  
Db 1546 CACAGAACCCAGGAGTACCCCATCA 1521  
RESULT 19  
ABV25077/c  
ID ABV25077 standard; cDNA; 2420 BP.  
XX AC ABV25077;  
XX 16-SEP-2002 (first entry)  
XX Human prostate expression marker cDNA 25068.  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX KW pharmacogenomic marker; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200160860-A2.  
XX 23-AUG-2001.  
XX 20-FEB-2001; 2001WO-US005171.  
PF

XX 17-FEB-2000; 2000US-0183319P.  
XX 16-MAR-2000; 2000US-0189862P.  
XX 25-MAY-2000; 2000US-0207454P.  
XX 09-JUN-2000; 2000US-0211314P.  
XX 18-JUL-2000; 2000US-0219007P.  
XX 13-DEC-2000; 2000US-0255281P.  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX Schlegel R, Endege WO, Monahan JE;  
XX WPI; 2001-662795/76.  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
XX PT prostate cells and correlating with presence of prostate cancer, useful  
XX PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX PS Claim 1; Page 4873-4874; 11750pp; English.  
XX The invention relates to an isolated nucleic acid molecule (I) comprising  
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
XX CC specification or its complement. (I) is useful for: (a) assessing whether  
XX CC a patient is afflicted with prostate cancer; (b) monitoring the  
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy  
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
XX CC determining whether prostate cancer has metastasized in a patient; (h)  
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient  
XX CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX SQ Sequence 2420 BP; 641 A; 458 C; 473 G; 838 T; 0 U; 10 Other;  
Query Match 65.3%; Score 19.6; DB 5; Length 2420;  
Best Local Similarity 84.6%; Pred. No. 2.8e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 4 CGCAGAACCCAGGAGTACCCCATGA 29  
Db 1546 CACAGAACCCAGGAGTACCCCATCA 1521  
RESULT 20  
ABL27011  
ID ABL27011 standard; DNA; 2427 BP.  
XX AC ABL27011;  
XX 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 32506.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical; gene; ds.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li FWD, Myers EW;  
XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 32506; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 2427 BP; 681 A; 694 C; 681 G; 371 T; 0 U; 0 Other;  
SQ  
Query Match 65.3%; Score 19.6; DB 4; Length 2427;  
Best Local Similarity 84.6%; Pred. No. 2.8e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 5 GCAGAACCCAGGAGTAGCCCAATGAG 30  
Db 2138 GCAGCACCAGGAGCAGCAACAATCAG 2163  
RESULT 21  
ABL27010  
ID ABL27010 standard; DNA; 6200 BP.  
XX  
XX ABL27010;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 32503.  
XX  
XX Drosophila; developmental biology; cell signaling; insecticide;  
KW pharmaceutical; gene; ds.  
KW  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 32503; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 6200 BP; 1834 A; 1436 C; 1429 G; 1501 T; 0 U; 0 Other;  
SQ  
Query Match 65.3%; Score 19.6; DB 4; Length 6200;  
Best Local Similarity 84.6%; Pred. No. 3.2e+02;  
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 5 GCAGAACCCAGGAGTAGCCCAATGAG 30  
Db 4454 GCAGCACCAGGAGCAGCAACAATCAG 4479  
RESULT 22  
ABA61254  
ID ABA61254 standard; DNA; 537 BP.  
XX  
XX ABA61254;  
XX  
XX 01-FEB-2002 (first entry)  
XX  
XX Human foetal liver single exon nucleic acid probe #9559.  
DE  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200157277-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US000669.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human foetal liver.  
XX  
XX Claim 1; SEQ ID NO 9559; 639pp + Sequence Listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human foetal liver. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;  
SQ  
Query Match 64.7%; Score 19.4; DB 4; Length 537;  
Best Local Similarity 79.3%; Pred. No. 2.7e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 2 GGCGCAGAACCCAGGAGTAGCCCAATGAG 30  
Db 49 GGCCAGAACCCAGGAGGCTCCAGTCAG 77

```
RESULT 23
AAI41155
ID AAI41155 standard; DNA; 537 BP.
XX
XX AAI41155;
XX
XX 17-OCT-2001 (first entry)
XX
XX DE Probe #9841 used to measure gene expression in human placenta sample.
XX
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX
XX PN WO200157272-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000663.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488897/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX
XX PS Claim 25; SEQ ID NO 9841; 654pp; English.
XX
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders
XX
XX SQ Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
XX
XX Query Match 64.7%; Score 19.4; DB 4; Length 537;
XX Best Local Similarity 79.3%; Pred. No. 2.7e+02;
XX Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX QY 2 GGCGCAGAACCCAGGAGTAGCACCACATGAG 30
XX ||||| ||||| ||||| ||||| |||||
XX Db 49 GGCCGAGAACCCAGGAGTAGCACCACATGAG 77
XX
XX RESULT 24
ABA29093
ID ABA29093 standard; DNA; 537 BP.
XX
XX AC ABA29093;
XX
XX DT 23-JAN-2002 (first entry)
XX
XX DE Probe #7559 for gene expression analysis in human heart cell sample.
XX
XX KW Human; gene expression; heart; microarray; vascular system; probe;
XX KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease; ss.
XX
XX OS Homo sapiens.
XX
```

```
PN WO200157274-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000666.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX DR WPI; 2001-488899/53.
XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT hearts.
XX
XX PS Claim 1; SEQ ID NO 7559; 530pp; English.
XX
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart. The
XX CC present sequence is one such probe. The probes may be used for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from the human heart via microarrays. By measuring gene expression, the
XX CC probes are useful for predicting, diagnosing, grading, staging,
XX CC monitoring and prognosing diseases of the human heart and vascular system
XX CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX CC congenital heart disease. Note: The sequence data for this patent did not
XX CC form part of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 537 BP; 99 A; 152 C; 177 G; 109 T; 0 U; 0 Other;
XX
XX Query Match 64.7%; Score 19.4; DB 4; Length 537;
XX Best Local Similarity 79.3%; Pred. No. 2.7e+02;
XX Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX QY 2 GGCGCAGAACCCAGGAGTAGCACCACATGAG 30
XX ||||| ||||| ||||| ||||| |||||
XX Db 49 GGCCGAGAACCCAGGAGTAGCACCACATGAG 77
XX
XX RESULT 25
AAK35442
ID AAK35442 standard; DNA; 537 BP.
XX
XX AC AAK35442;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 9999.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000668.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
```







Job time : 187.506 secs

XX The present invention describes a computer readable medium which has the  
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
CC on it, or a representative fragment or a sequence at least 95% identical  
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
CC pneumoniae. The present invention also describes an isolated nucleic acid  
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae  
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
CC by a process comprising: (a) screening a genomic DNA library using as a  
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
CC 391, identifying members of the library which contain sequences that  
CC hybridize to the target sequence and isolating the nucleic acid molecules  
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is  
CC homologous to amplification primers derived from the fragment of the S.  
CC pneumoniae genome to prime the amplification and isolating the amplified  
CC sequences. The computer readable medium can be used in a computer-based  
CC system for identifying fragments of the S. pneumoniae genome of  
CC commercial importance, or expression modulating fragments of the S.  
CC pneumoniae genome. Products from the present invention can be used in  
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines  
CC for S. pneumoniae

XX Sequence 21706 BP; 6489 A; 3945 C; 4879 G; 6392 T; 0 U; 0 Other;

Query Match 64.0%; Score 19.2; DB 2; Length 21706;  
Best Local Similarity 87.5%; Pred. No. 5.6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAGAACCCAGGAGTAGCCCAATG 28  
DB 6961 GAAGAACCAGAGTAGCTCCAATG 6984

## RESULT 30

ABS56454\_14/C

Continuation (15 of 22) of ABS56454 from base 1400001 (Streptococcus pneumoniae type 4 S

WP Sequence split into 22 fragments LOCUS ABS56454 Accession ABS56454

WP Fragment Name Begin End

|                |         |         |
|----------------|---------|---------|
| WP ABS56454_00 | 1       | 110000  |
| WP ABS56454_01 | 100001  | 210000  |
| WP ABS56454_02 | 200001  | 310000  |
| WP ABS56454_03 | 300001  | 410000  |
| WP ABS56454_04 | 400001  | 510000  |
| WP ABS56454_05 | 500001  | 610000  |
| WP ABS56454_06 | 600001  | 710000  |
| WP ABS56454_07 | 700001  | 810000  |
| WP ABS56454_08 | 800001  | 910000  |
| WP ABS56454_09 | 900001  | 1010000 |
| WP ABS56454_10 | 1000001 | 1110000 |
| WP ABS56454_11 | 1100001 | 1210000 |
| WP ABS56454_12 | 1200001 | 1310000 |
| WP ABS56454_13 | 1300001 | 1410000 |
| WP ABS56454_14 | 1400001 | 1510000 |
| WP ABS56454_15 | 1500001 | 1610000 |
| WP ABS56454_16 | 1600001 | 1710000 |
| WP ABS56454_17 | 1700001 | 1810000 |
| WP ABS56454_18 | 1800001 | 1910000 |
| WP ABS56454_19 | 1900001 | 2010000 |
| WP ABS56454_20 | 2000001 | 2110000 |
| WP ABS56454_21 | 2100001 | 2162598 |

Query Match 64.0%; Score 19.2; DB 7; Length 110000;  
Best Local Similarity 87.5%; Pred. No. 7.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAGAACCCAGGAGTAGCCCAATG 28  
DB 56073 GAAGAACCAGAGTAGCTCCAATG 56050

Search completed: June 20, 2004, 10:18:21



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:48:37 ; Search time 1401.9 Seconds  
(without alignments)  
639.034 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcggaaccagtagtagcaccatgag 30

Scoring table: IDENTITY\_NUC

Gapcp 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

EST:

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estnu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_hcc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_hcc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_pln.\*

20: em\_gss\_vrt.\*

21: em\_gss\_fun.\*

22: em\_gss\_mam.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_rod.\*

26: em\_gss\_phg.\*

27: em\_gss\_vrl.\*

28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID          | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 21.2  | 70.7        | 670    | 13 BY763514 | BY763514 BY763514  |
| 2          | 21.2  | 70.7        | 691    | 14 CB551691 | CB551691 MMSPP0057 |
| 3          | 20.8  | 69.3        | 298    | 10 BB025068 | BB025068 BB025068  |
| 4          | 20.8  | 69.3        | 839    | 14 CF593587 | CF593587 AGENCOURT |

|          |           |     |    |          |     |      |      |   |    |
|----------|-----------|-----|----|----------|-----|------|------|---|----|
| CK130228 | AGENCOURT | 767 | 14 | CK130228 | 767 | 68.7 | 20.6 | c | 5  |
| CK130232 | AGENCOURT | 767 | 14 | CK130232 | 767 | 68.7 | 20.6 | c | 6  |
| CF271960 | AGENCOURT | 768 | 14 | CF271960 | 768 | 68.7 | 20.6 | c | 7  |
| CK130229 | AGENCOURT | 771 | 14 | CK130229 | 771 | 68.7 | 20.6 | c | 8  |
| CK130230 | AGENCOURT | 779 | 14 | CK130230 | 779 | 68.7 | 20.6 | c | 9  |
| CF271962 | AGENCOURT | 795 | 14 | CF271962 | 795 | 68.7 | 20.6 | c | 10 |
| CK130231 | AGENCOURT | 804 | 14 | CK130231 | 804 | 68.7 | 20.6 | c | 11 |
| BZ773303 | mcv67d11. | 986 | 28 | BZ773303 | 986 | 68.7 | 20.6 | c | 12 |
| BF444251 | 262053 MA | 227 | 10 | BF444251 | 227 | 68.0 | 20.4 | c | 13 |
| BQ136744 | NF002D06S | 230 | 13 | BQ136744 | 230 | 68.0 | 20.4 | c | 14 |
| BZ246711 | BB246711  | 241 | 10 | BB246711 | 241 | 68.0 | 20.4 | c | 15 |
| BE318013 | NF061H03L | 250 | 10 | BE318013 | 250 | 68.0 | 20.4 | c | 16 |
| BE316071 | NF031B01L | 281 | 10 | BE316071 | 281 | 68.0 | 20.4 | c | 17 |
| BF404844 | NF037A02I | 290 | 10 | BF404844 | 290 | 68.0 | 20.4 | c | 18 |
| BI266129 | NF100G10I | 293 | 12 | BI266129 | 293 | 68.0 | 20.4 | c | 19 |
| BI267471 | NF1D5E03I | 312 | 12 | BI267471 | 312 | 68.0 | 20.4 | c | 20 |
| BI265945 | NF092F12I | 319 | 12 | BI265945 | 319 | 68.0 | 20.4 | c | 21 |
| AW127696 | M110445 D | 349 | 10 | AW127696 | 349 | 68.0 | 20.4 | c | 22 |
| BE323556 | NF01CH11P | 358 | 10 | BE323556 | 358 | 68.0 | 20.4 | c | 23 |
| BE315922 | NF098C04L | 361 | 10 | BE315922 | 361 | 68.0 | 20.4 | c | 24 |
| AW683712 | NF018A04L | 384 | 10 | AW683712 | 384 | 68.0 | 20.4 | c | 25 |
| BE324258 | NF020A04P | 382 | 10 | BE324258 | 382 | 68.0 | 20.4 | c | 26 |
| BE322941 | NF025C04I | 386 | 10 | BE322941 | 386 | 68.0 | 20.4 | c | 27 |
| BE323559 | NF003F04P | 386 | 10 | BE323559 | 386 | 68.0 | 20.4 | c | 28 |
| BE249756 | NF021H07L | 389 | 10 | BE249756 | 389 | 68.0 | 20.4 | c | 29 |
| BI263513 | NF087A01P | 398 | 12 | BI263513 | 398 | 68.0 | 20.4 | c | 30 |
| AW911625 | U188B03.Y | 409 | 10 | AW911625 | 409 | 68.0 | 20.4 | c | 31 |
| BF005193 | EST43369I | 425 | 10 | BF005193 | 425 | 68.0 | 20.4 | c | 32 |
| BE318665 | NF040G06L | 438 | 10 | BE318665 | 438 | 68.0 | 20.4 | c | 33 |
| BI265760 | NF086E05I | 447 | 12 | BI265760 | 447 | 68.0 | 20.4 | c | 34 |
| BE317431 | NF069C12L | 458 | 10 | BE317431 | 458 | 68.0 | 20.4 | c | 35 |
| BF464422 | U1-N-CGDP | 462 | 10 | BF464422 | 462 | 68.0 | 20.4 | c | 36 |
| BE318612 | NF042C05L | 473 | 10 | BE318612 | 473 | 68.0 | 20.4 | c | 37 |
| BE324767 | NF017C08P | 505 | 10 | BE324767 | 505 | 68.0 | 20.4 | c | 38 |
| BF640358 | NF031H09I | 507 | 10 | BF640358 | 507 | 68.0 | 20.4 | c | 39 |
| BF520833 | EST458306 | 513 | 10 | BF520833 | 513 | 68.0 | 20.4 | c | 40 |
| BF448827 | NF002B06I | 523 | 12 | BF448827 | 523 | 68.0 | 20.4 | c | 41 |
| BE316819 | NF066D08L | 526 | 10 | BE316819 | 526 | 68.0 | 20.4 | c | 42 |
| BI264512 | NF108E04P | 528 | 12 | BI264512 | 528 | 68.0 | 20.4 | c | 43 |
| BE318717 | NF042D05L | 539 | 10 | BE318717 | 539 | 68.0 | 20.4 | c | 44 |
| BQ139956 | NF027B08P | 539 | 13 | BQ139956 | 539 | 68.0 | 20.4 | c | 45 |
| BF637160 | NF049H04L | 543 | 10 | BF637160 | 543 | 68.0 | 20.4 | c | 46 |
| BI263324 | NF088H07P | 557 | 12 | BI263324 | 557 | 68.0 | 20.4 | c | 47 |
| BE124199 | EST394334 | 558 | 10 | BE124199 | 558 | 68.0 | 20.4 | c | 48 |
| BF640661 | NF044D10I | 561 | 10 | BF640661 | 561 | 68.0 | 20.4 | c | 49 |
| BE319051 | NF034D11L | 584 | 10 | BE319051 | 584 | 68.0 | 20.4 | c | 50 |
| BF639376 | NF011F04I | 570 | 10 | BF639376 | 570 | 68.0 | 20.4 | c | 51 |
| AW683744 | NF018D07L | 576 | 10 | AW683744 | 576 | 68.0 | 20.4 | c | 52 |
| BF638168 | NF044G01P | 577 | 10 | BF638168 | 577 | 68.0 | 20.4 | c | 53 |
| BF006139 | EST434709 | 577 | 10 | BF006139 | 577 | 68.0 | 20.4 | c | 54 |
| BF004897 | EST433458 | 581 | 10 | BF004897 | 581 | 68.0 | 20.4 | c | 55 |
| BE316701 | NF065A06L | 582 | 10 | BE316701 | 582 | 68.0 | 20.4 | c | 56 |
| BF639007 | NF094B12P | 583 | 10 | BF639007 | 583 | 68.0 | 20.4 | c | 57 |
| BF637828 | NF043D08P | 587 | 10 | BF637828 | 587 | 68.0 | 20.4 | c | 58 |
| AW692465 | NF051G09S | 590 | 10 | AW692465 | 590 | 68.0 | 20.4 | c | 59 |
| BF323679 | NF005A11P | 595 | 10 | BF323679 | 595 | 68.0 | 20.4 | c | 60 |
| BF005033 | EST433531 | 597 | 10 | BF005033 | 597 | 68.0 | 20.4 | c | 61 |
| AW980360 | EST391513 | 602 | 10 | AW980360 | 602 | 68.0 | 20.4 | c | 62 |
| BE324591 | NF014E04P | 604 | 10 | BE324591 | 604 | 68.0 | 20.4 | c | 63 |
| BI268290 | NF120D11I | 605 | 12 | BI268290 | 605 | 68.0 | 20.4 | c | 64 |
| BF642786 | NF072F04I | 608 | 10 | BF642786 | 608 | 68.0 | 20.4 | c | 65 |
| BG457889 | NF033B03P | 617 | 12 | BG457889 | 617 | 68.0 | 20.4 | c | 66 |
| AW683772 | NF018G11L | 621 | 10 | AW683772 | 621 | 68.0 | 20.4 | c | 67 |
| BE324010 | NF012D11P | 623 | 10 | BE324010 | 623 | 68.0 | 20.4 | c | 68 |
| BQ138571 | NF004G07P | 630 | 13 | BQ138571 | 630 | 68.0 | 20.4 | c | 69 |
| AW256790 | EST304927 | 631 | 10 | AW256790 | 631 | 68.0 | 20.4 | c | 70 |
| BG456877 | NF088F10P | 633 | 12 | BG456877 | 633 | 68.0 | 20.4 | c | 71 |
| BM614057 | EST592150 | 634 | 12 | BM614057 | 634 | 68.0 | 20.4 | c | 72 |
| BQ138282 | NF001E06P | 635 | 13 | BQ138282 | 635 | 68.0 | 20.4 | c | 73 |
| BG456673 | NF082E07P | 639 | 12 | BG456673 | 639 | 68.0 | 20.4 | c | 74 |
| BF520572 | EST458045 | 640 | 10 | BF520572 | 640 | 68.0 | 20.4 | c | 75 |

|       |      |      |     |    |          |           |       |      |      |      |    |          |                    |
|-------|------|------|-----|----|----------|-----------|-------|------|------|------|----|----------|--------------------|
| C 78  | 20.4 | 68.0 | 644 | 12 | BQ456201 | NR075A02P | 151   | 19.4 | 64.7 | 289  | 13 | BQ305306 | NR0-BT200          |
| C 79  | 20.4 | 68.0 | 645 | 10 | BE317057 | NR063D03L | C 152 | 19.4 | 64.7 | 291  | 9  | AA344653 | EST50642           |
| C 80  | 20.4 | 68.0 | 646 | 13 | BQ138629 | NR005C04P | 153   | 19.4 | 64.7 | 305  | 10 | BE655367 | UI-M-BHO-          |
| C 81  | 20.4 | 68.0 | 647 | 10 | B249506  | NR021A04L | C 154 | 19.4 | 64.7 | 309  | 12 | BG559599 | RHI22-69           |
| C 82  | 20.4 | 68.0 | 648 | 10 | B2637864 | NR043F11P | C 155 | 19.4 | 64.7 | 328  | 29 | CG707991 | 1119004PI          |
| C 83  | 20.4 | 68.0 | 649 | 10 | B2637864 | NR043F11P | C 156 | 19.4 | 64.7 | 330  | 9  | AI426118 | mb43804.x          |
| C 84  | 20.4 | 68.0 | 650 | 10 | B2637864 | NR043F11P | C 157 | 19.4 | 64.7 | 332  | 13 | BY173156 | BV173156           |
| C 85  | 20.4 | 68.0 | 651 | 12 | B2637864 | NR043F11P | C 158 | 19.4 | 64.7 | 336  | 29 | CG707847 | 1119004A1          |
| C 86  | 20.4 | 68.0 | 652 | 12 | B2637864 | NR043F11P | C 159 | 19.4 | 64.7 | 338  | 9  | AA188722 | 2P78503.r          |
| C 87  | 20.4 | 68.0 | 653 | 12 | B2637864 | NR043F11P | C 160 | 19.4 | 64.7 | 339  | 14 | W82305   | W82305 mf0212.r    |
| C 88  | 20.4 | 68.0 | 654 | 12 | B2637864 | NR043F11P | C 161 | 19.4 | 64.7 | 411  | 9  | AI182795 | uc65411.r          |
| C 89  | 20.4 | 68.0 | 655 | 12 | B2637864 | NR043F11P | C 162 | 19.4 | 64.7 | 422  | 13 | BY158314 | BY158314           |
| C 90  | 20.4 | 68.0 | 656 | 12 | B2637864 | NR043F11P | C 163 | 19.4 | 64.7 | 426  | 13 | BY542403 | BY542403           |
| C 91  | 20.4 | 68.0 | 657 | 12 | B2637864 | NR043F11P | C 164 | 19.4 | 64.7 | 429  | 9  | AA968188 | uh14610.r          |
| C 92  | 20.4 | 68.0 | 658 | 12 | B2637864 | NR043F11P | C 165 | 19.4 | 64.7 | 430  | 13 | BY551293 | BY551293           |
| C 93  | 20.4 | 68.0 | 659 | 12 | B2637864 | NR043F11P | C 166 | 19.4 | 64.7 | 432  | 14 | CB820248 | EST 1240           |
| C 94  | 20.4 | 68.0 | 660 | 12 | B2637864 | NR043F11P | C 167 | 19.4 | 64.7 | 437  | 9  | AA671087 | AA671087 VK22411.r |
| C 95  | 20.4 | 68.0 | 661 | 12 | B2637864 | NR043F11P | C 168 | 19.4 | 64.7 | 438  | 10 | BB825444 | BB825444           |
| C 96  | 20.4 | 68.0 | 662 | 12 | B2637864 | NR043F11P | C 169 | 19.4 | 64.7 | 444  | 9  | AA656665 | V-49F08.s          |
| C 97  | 20.4 | 68.0 | 663 | 12 | B2637864 | NR043F11P | C 170 | 19.4 | 64.7 | 452  | 13 | BY545446 | BY545446           |
| C 98  | 20.4 | 68.0 | 664 | 12 | B2637864 | NR043F11P | C 171 | 19.4 | 64.7 | 453  | 10 | AW908257 | BY552318           |
| C 99  | 20.4 | 68.0 | 665 | 12 | B2637864 | NR043F11P | C 172 | 19.4 | 64.7 | 458  | 28 | AQ335940 | BS52318            |
| C 100 | 20.4 | 68.0 | 666 | 12 | B2637864 | NR043F11P | C 173 | 19.4 | 64.7 | 463  | 13 | BQ297619 | BS52318            |
| C 101 | 20.4 | 68.0 | 667 | 12 | B2637864 | NR043F11P | C 174 | 19.4 | 64.7 | 468  | 10 | BB823361 | BB823361           |
| C 102 | 20.4 | 68.0 | 668 | 12 | B2637864 | NR043F11P | C 175 | 19.4 | 64.7 | 465  | 10 | BB823361 | BB823361           |
| C 103 | 20.4 | 68.0 | 669 | 12 | B2637864 | NR043F11P | C 176 | 19.4 | 64.7 | 465  | 10 | BB823361 | BB823361           |
| C 104 | 20.4 | 68.0 | 670 | 12 | B2637864 | NR043F11P | C 177 | 19.4 | 64.7 | 468  | 13 | BY598882 | BY598882           |
| C 105 | 20.4 | 68.0 | 671 | 12 | B2637864 | NR043F11P | C 178 | 19.4 | 64.7 | 479  | 13 | BY448067 | BY448067           |
| C 106 | 20.4 | 68.0 | 672 | 12 | B2637864 | NR043F11P | C 179 | 19.4 | 64.7 | 479  | 13 | BY689321 | BY689321           |
| C 107 | 20.4 | 68.0 | 673 | 12 | B2637864 | NR043F11P | C 180 | 19.4 | 64.7 | 488  | 14 | N74643   | za55101.sl         |
| C 108 | 20.4 | 68.0 | 674 | 12 | B2637864 | NR043F11P | C 181 | 19.4 | 64.7 | 500  | 28 | AZ179652 | SP-0165.B          |
| C 109 | 20.4 | 68.0 | 675 | 12 | B2637864 | NR043F11P | C 182 | 19.4 | 64.7 | 509  | 14 | BD335083 | STrPu536.          |
| C 110 | 20.4 | 68.0 | 676 | 12 | B2637864 | NR043F11P | C 183 | 19.4 | 64.7 | 513  | 28 | BD335083 | BD335083           |
| C 111 | 20.4 | 68.0 | 677 | 12 | B2637864 | NR043F11P | C 184 | 19.4 | 64.7 | 519  | 29 | CE395685 | tigt-gss-          |
| C 112 | 20.4 | 68.0 | 678 | 12 | B2637864 | NR043F11P | C 185 | 19.4 | 64.7 | 531  | 10 | CG726707 | 111909B1           |
| C 113 | 20.4 | 68.0 | 679 | 12 | B2637864 | NR043F11P | C 186 | 19.4 | 64.7 | 536  | 29 | CG726707 | 111909B1           |
| C 114 | 20.4 | 68.0 | 680 | 12 | B2637864 | NR043F11P | C 187 | 19.4 | 64.7 | 556  | 28 | AZ747080 | RPC1-24-7          |
| C 115 | 20.4 | 68.0 | 681 | 12 | B2637864 | NR043F11P | C 188 | 19.4 | 64.7 | 583  | 28 | AZ758572 | IM050F09           |
| C 116 | 20.4 | 68.0 | 682 | 12 | B2637864 | NR043F11P | C 189 | 19.4 | 64.7 | 585  | 10 | BE402590 | CSB009E08          |
| C 117 | 20.4 | 68.0 | 683 | 12 | B2637864 | NR043F11P | C 190 | 19.4 | 64.7 | 585  | 13 | BQ608170 | BRY 4072           |
| C 118 | 20.4 | 68.0 | 684 | 12 | B2637864 | NR043F11P | C 191 | 19.4 | 64.7 | 601  | 9  | AI981752 | pat-Pk006          |
| C 119 | 20.4 | 68.0 | 685 | 12 | B2637864 | NR043F11P | C 192 | 19.4 | 64.7 | 633  | 14 | CD776569 | UI-N-A00-          |
| C 120 | 20.4 | 68.0 | 686 | 12 | B2637864 | NR043F11P | C 193 | 19.4 | 64.7 | 634  | 10 | BE368075 | 60121971           |
| C 121 | 20.4 | 68.0 | 687 | 12 | B2637864 | NR043F11P | C 194 | 19.4 | 64.7 | 651  | 12 | BI295028 | UI-R-DK0-          |
| C 122 | 20.4 | 68.0 | 688 | 12 | B2637864 | NR043F11P | C 195 | 19.4 | 64.7 | 672  | 12 | BI295028 | UI-R-DK0-          |
| C 123 | 20.4 | 68.0 | 689 | 12 | B2637864 | NR043F11P | C 196 | 19.4 | 64.7 | 678  | 10 | BB506736 | BS506736           |
| C 124 | 20.4 | 68.0 | 690 | 12 | B2637864 | NR043F11P | C 197 | 19.4 | 64.7 | 683  | 28 | BB506736 | BS506736           |
| C 125 | 20.4 | 68.0 | 691 | 12 | B2637864 | NR043F11P | C 198 | 19.4 | 64.7 | 693  | 12 | BI961871 | MON01-7.F          |
| C 126 | 20.4 | 68.0 | 692 | 12 | B2637864 | NR043F11P | C 199 | 19.4 | 64.7 | 699  | 12 | BG974597 | 602841831          |
| C 127 | 20.4 | 68.0 | 693 | 12 | B2637864 | NR043F11P | C 200 | 19.4 | 64.7 | 702  | 29 | CG621439 | CG621439           |
| C 128 | 20.4 | 68.0 | 694 | 12 | B2637864 | NR043F11P | C 201 | 19.4 | 64.7 | 702  | 29 | CG621439 | CG621439           |
| C 129 | 20.4 | 68.0 | 695 | 12 | B2637864 | NR043F11P | C 202 | 19.4 | 64.7 | 704  | 29 | CG575967 | CG575967           |
| C 130 | 20.4 | 68.0 | 696 | 12 | B2637864 | NR043F11P | C 203 | 19.4 | 64.7 | 711  | 29 | CG588083 | CG588083           |
| C 131 | 20.4 | 68.0 | 697 | 12 | B2637864 | NR043F11P | C 204 | 19.4 | 64.7 | 734  | 28 | BZ526088 | OGALW35TM          |
| C 132 | 20.4 | 68.0 | 698 | 12 | B2637864 | NR043F11P | C 205 | 19.4 | 64.7 | 734  | 28 | BZ526088 | OGALW35TM          |
| C 133 | 20.4 | 68.0 | 699 | 12 | B2637864 | NR043F11P | C 206 | 19.4 | 64.7 | 796  | 10 | BE034896 | tigt-gss-          |
| C 134 | 20.4 | 68.0 | 700 | 12 | B2637864 | NR043F11P | C 207 | 19.4 | 64.7 | 804  | 12 | BG539731 | 602867702          |
| C 135 | 20.4 | 68.0 | 701 | 12 | B2637864 | NR043F11P | C 208 | 19.4 | 64.7 | 805  | 29 | CG558304 | OG0A9090TV         |
| C 136 | 20.4 | 68.0 | 702 | 12 | B2637864 | NR043F11P | C 209 | 19.4 | 64.7 | 808  | 14 | CB819759 | EST 751.G          |
| C 137 | 20.4 | 68.0 | 703 | 12 | B2637864 | NR043F11P | C 210 | 19.4 | 64.7 | 810  | 28 | BZ726232 | OGEDC38TC          |
| C 138 | 20.4 | 68.0 | 704 | 12 | B2637864 | NR043F11P | C 211 | 19.4 | 64.7 | 821  | 29 | CG610468 | OGUB066TV          |
| C 139 | 20.4 | 68.0 | 705 | 12 | B2637864 | NR043F11P | C 212 | 19.4 | 64.7 | 821  | 29 | CG610468 | OGUB066TV          |
| C 140 | 20.4 | 68.0 | 706 | 12 | B2637864 | NR043F11P | C 213 | 19.4 | 64.7 | 839  | 29 | CG345029 | OG5A34TC           |
| C 141 | 20.4 | 68.0 | 707 | 12 | B2637864 | NR043F11P | C 214 | 19.4 | 64.7 | 849  | 28 | CG345029 | OG5A34TC           |
| C 142 | 20.4 | 68.0 | 708 | 12 | B2637864 | NR043F11P | C 215 | 19.4 | 64.7 | 849  | 28 | CG345029 | OG5A34TC           |
| C 143 | 20.4 | 68.0 | 709 | 12 | B2637864 | NR043F11P | C 216 | 19.4 | 64.7 | 849  | 28 | CG345029 | OG5A34TC           |
| C 144 | 20.4 | 68.0 | 710 | 12 | B2637864 | NR043F11P | C 217 | 19.4 | 64.7 | 893  | 29 | CF215604 | CAST0002           |
| C 145 | 20.4 | 68.0 | 711 | 12 | B2637864 | NR043F11P | C 218 | 19.4 | 64.7 | 930  | 14 | CF215604 | CAST0002           |
| C 146 | 20.4 | 68.0 | 712 | 12 | B2637864 | NR043F11P | C 219 | 19.4 | 64.7 | 940  | 14 | CG993468 | AGENCOURT          |
| C 147 | 20.4 | 68.0 | 713 | 12 | B2637864 | NR043F11P | C 220 | 19.4 | 64.7 | 964  | 29 | CG325217 | OG3BC91TV          |
| C 148 | 20.4 | 68.0 | 714 | 12 | B2637864 | NR043F11P | C 221 | 19.4 | 64.7 | 1111 | 13 | BUI15986 | 603136432          |
| C 149 | 20.4 | 68.0 | 715 | 12 | B2637864 | NR043F11P | C 222 | 19.4 | 64.7 | 1265 | 11 | AK052797 | Mus muscu          |
| C 150 | 20.4 | 68.0 | 716 | 12 | B2637864 | NR043F11P | C 223 | 19.4 | 64.7 | 1633 | 11 | AK015988 | Mus muscu          |



Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose

assistance we gratefully acknowledge.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
 source

1. .670  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="NOD"  
 /db\_xref="taxon:10090"  
 /clone="E430037109"  
 /tissue\_type="thymus"  
 /cell\_type="thymic cells"  
 /clone\_lib="RIKEN full-length enriched, 2 days neonate thymus thymic cells (NOD)"

## ORIGIN

Query Match 70.7%; Score 21.2; DB 13; Length 670;  
 Best Local Similarity 88.5%; Pred. No. 2.7e+03;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCGCGAGAACCGAGGTAGCACCMA 26  
 |||||  
 DB 108 TGCGCGAGAGCAGGAGTGCTCCAA 133

## RESULT 2.

CB551691/c 691 bp mRNA linear EST 01-JUN-2003

LOCUS MWSP0057 H06 WMSP Macaca mulatta cDNA, mRNA sequence.  
 DEFINITION

ACCESSION CB551691

VERSION CB551691.1 GI:31300886

KEYWORDS EST.

SOURCE Macaca mulatta (rhesus monkey)

ORGANISM Macaca mulatta

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;

Cercopitheciinae; Macaca.

1 (bases 1 to 691)

Katze, M.G., Bungarner, R., Korth, M., Feldman, R., Amjadi, M. and

Holzman, I.

Expressed sequence tags from Rhesus macaque spleen

Unpublished (2002)

Contact: Holzman T

Katze Lab

University of Washington

Box 358070, Seattle, WA 98195-8070, USA

Tel: 206 732 6156

Fax: 206 732 6055

Email: ted@locke.hs.washington.edu

Similar to GenBank entry AK091287 AK091287 Homo sapiens cDNA

FLJ33968 fls, clone DFNES2001096. 7/2002

Plate: MWSP0057 row: H column: 06.

FEATURES  
 source  
 Location/Qualifiers  
 1. .691  
 /organism="Macaca mulatta"  
 /mol\_type="mRNA"

/db\_xref="taxon:9544"  
 /sex="male"  
 /cell\_type="monocytes"  
 /dev\_stage="adult"  
 /clone\_lib="WMSP"  
 /notes="Organ: spleen"

## ORIGIN

Query Match 70.7%; Score 21.2; DB 14; Length 691;  
 Best Local Similarity 88.5%; Pred. No. 2.7e+03;  
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CGCAGAACCGAGGTAGCACCACATGA 29  
 |||||  
 DB 424 CGCAGAACCGAGGTAGCACCACATCA 399

## RESULT 3

BB025068 RIKEN full-length enriched, adult male pituitary gland Mus  
 musculus cDNA clone 5330426A17 3' similar to AJ222586 Mus musculus  
 mRNA for NEFA protein, mRNA sequence.

BB025068 298 bp mRNA linear EST 23-JUN-2000  
 LOCUS BB025068

DEFINITION

ACCESSION

VERSION BB025068.1 GI:8199385

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 298)

Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,

Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H.,

Okazaki, Y., Omo, T., Owa, C., Saito, H., Sakai, C., Sato, K.,

Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T.,

Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,

Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A.,

Watanabe, S., Yamamura, T., Yamahata, I., Yano, R., Yasunishi, A.,

Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and

Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gs.c.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Sasaki, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,

Sasai, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

19-44 (1999)

Itoh, M., Kitsumi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,

Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

further details.

FEATURES

source

Location/Qualifiers

1. .298

/organism="Mus musculus"

FEATURES  
source

of the pDNR-Dual vector. Library constructed by Dr. Narayan Bhat, Earl Bere III and Hongling Liao (Gene Expression Laboratory, Research Technology Program, SAIC Frederick, NCI-Frederick, Frederick, MD 21702). For information on which gene each clone represents, please visit our anonymous ftp site at  
ftp://image.llnl.gov/image/rearrayed\_plates/IRBK.presv.dat  
a Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 68.7%; Score 20.6; DB 14; Length 767;  
Best Local Similarity 85.2%; Pred. No. 4.6e+03;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
|||||  
Db 72 TGGCGCTGAACCCAGGAGTAGCAGAGT 46  
|||||

## RESULT 6

CK130232/c  
LOCUS CK130232.1 767 bp mRNA linear EST 02-DEC-2003  
DEFINITION AGENCOURT 15196775 NIH\_MGC.195 Homo sapiens cDNA clone  
IMAGE:7002138 5', mRNA sequence.

ACCESSION CK130232

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 767)

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Narayan Bhat  
cDNA Library Preparation: Bhat Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: IRBK4 row: b column: 04

High quality sequence start: 14

High quality sequence stop: 464.

Location/Qualifiers

1. .767

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:7002138"

/tissue\_type="mixed"

/lab\_host="DH5A (TI phage-resistant)"

/clone\_lib="NIH MGC 195"

/note="Vector: pDNR-Dual; Site 1: loxp-Sall; Site 2:

loxp-HindIII; Clones from this library have been

PCR-amplified using gene-specific primers to contain the

complete open reading frame (based on known gene sequences

available from NCBI's RefSeq). Template for PCR is cDNA

derived from either pooled cytoplasmic polyA RNA from 30

cells lines or pooled total RNA from 10 different tissues

(from BD Biosciences/Clontech and Washington University).

PCR products are directionally cloned into the loxp sites

of the pDNR-Dual vector. Library constructed by Dr.

Narayan Bhat, Earl Bere III and Hongling Liao (Gene

Expression Laboratory, Research Technology Program, SAIC

Frederick, NCI-Frederick, Frederick, MD 21702). For

information on which gene each clone represents, please

visit our anonymous ftp site at  
ftp://image.llnl.gov/image/rearrayed\_plates/IRBK.presv.dat  
a Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 68.7%; Score 20.6; DB 14; Length 767;  
Best Local Similarity 85.2%; Pred. No. 4.6e+03;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
|||||  
Db 107 TGGCGCTGAACCCAGGAGTAGCAGAGT 81  
|||||

## RESULT 7

CF271960/c

LOCUS CF271960

DEFINITION AGENCOURT 15196903 NIH\_MGC.195 Homo sapiens cDNA clone

IMAGE:7002142 5', mRNA sequence.

ACCESSION CF271960

VERSION CF271960.2

KEYWORDS GI:38559850

SOURCE EST.

ORGANISM Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 768)

AUTHORS NIH-MGC http://mgi.nci.nih.gov/

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

COMMENT On Aug 12, 2003 this sequence version replaced gi:33627872.

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Narayan Bhat

cDNA Library Preparation: Bhat Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: IRBK4 row: b column: 08

High quality sequence stop: 702.

Location/Qualifiers

1. .768

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:7002142"

/tissue\_type="mixed"

/lab\_host="DH5A (TI phage-resistant)"

/clone\_lib="NIH MGC 195"

/note="Vector: pDNR-Dual; Site 1: loxp-Sall; Site 2:

loxp-HindIII; Clones from this library have been

PCR-amplified using gene-specific primers to contain the

complete open reading frame (based on known gene sequences

available from NCBI's RefSeq). Template for PCR is cDNA

derived from either pooled cytoplasmic polyA RNA from 30

cells lines or pooled total RNA from 10 different tissues

(from BD Biosciences/Clontech and Washington University).

PCR products are directionally cloned into the loxp sites

of the pDNR-Dual vector. Library constructed by Dr.

Narayan Bhat, Earl Bere III and Hongling Liao (Gene

Expression Laboratory, Research Technology Program, SAIC

Frederick, NCI-Frederick, Frederick, MD 21702). For

information on which gene each clone represents, please

visit our anonymous ftp site at

ftp://image.llnl.gov/image/rearrayed\_plates/IRBK.presv.dat

a Note: this is a NIH\_MGC Library."

ORIGIN





```

DEFINITION  AGENCOURT 15196743 NIH MGC 195 Homo sapiens cDNA clone
IMAGE:7002137 5', mRNA sequence.
ACCESSION  CF271962
VERSION    CF271962.2 GI:38559851
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1 (bases 1 to 785)
AUTHORS  NIH-MGC http://mgs.nci.nih.gov/
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  On Aug 12, 2003 this sequence version replaced gi:33627874.
          Contact: Daniela S. Gerhard, Ph.D.
          Office of Cancer Genomics
          National Cancer Institute / NIH
          Bldg 31 Rm10A07 Bethesda, MD 20892
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Narayan Bhat
          cDNA Library Preparation: Bhat Laboratory
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: IRBK4 row: b column: 03
          High quality sequence start: 4
          High quality sequence stop: 764.

FEATURES             source
     source
     1..785
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:7002137"
         /tissue_type="mixed"
         /lab_host="DH5A (T1 phage-resistant)"
         /clone_lib="NIH MGC 195"
         /note="Vector: pDNR-Dual; Site 1: loxp-SalI; Site 2:
         loxp-HindIII; Clones from this library have been
         PCR-amplified using gene-specific primers to contain the
         complete open reading frame (based on known gene sequences
         available from NCBI's RefSeq). Template for PCR is cDNA
         derived from either pooled cytoplasmic polyA RNA from 30
         cells lines or pooled total RNA from 10 different tissues
         (from BD Biosciences/Clontech and Washington University).
         PCR products are directionally cloned into the loxp sites
         of the pDNR-Dual vector. Library constructed by Dr.
         Narayan Bhat, Earl Bere III and Hongling Liao (Gene
         Expression Laboratory, Research Technology Program, SAIC
         Frederick, NCI-Frederick, Frederick, MD 21702). For
         information on which gene each clone represents, please
         visit our anonymous ftp site at
         ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
         a Note: this is a NIH_MGC Library."

ORIGIN
Query Match      68.7%; Score 20.6; DB 14; Length 785;
Best Local Similarity 85.2%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCGCGAGAACCGAGGTAGCACCACAT 27
    |||||
DB 98 TGCGCGTGAACCGAGGTAGCAGAGT 72
    |||||

RESULT 11
CK130231/c
LOCUS
DEFINITION  AGENCOURT 15196807 NIH MGC 195 Homo sapiens cDNA clone
IMAGE:7002139 5', mRNA sequence.
ACCESSION  CK130231
VERSION    CK130231.1 GI:38622167

DEFINITION  AGENCOURT 15196743 NIH MGC 195 Homo sapiens cDNA clone
IMAGE:7002137 5', mRNA sequence.
ACCESSION  CF271962
VERSION    CF271962.2 GI:38559851
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1 (bases 1 to 804)
AUTHORS  NIH-MGC http://mgs.nci.nih.gov/
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  On Aug 12, 2003 this sequence version replaced gi:33627874.
          Contact: Daniela S. Gerhard, Ph.D.
          Office of Cancer Genomics
          National Cancer Institute / NIH
          Bldg 31 Rm10A07 Bethesda, MD 20892
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Narayan Bhat
          cDNA Library Preparation: Bhat Laboratory
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: IRBK4 row: b column: 05
          High quality sequence start: 3
          High quality sequence stop: 737.

FEATURES             source
     source
     1..804
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:7002139"
         /tissue_type="mixed"
         /lab_host="DH5A (T1 phage-resistant)"
         /clone_lib="NIH MGC 195"
         /note="Vector: pDNR-Dual; Site 1: loxp-SalI; Site 2:
         loxp-HindIII; Clones from this library have been
         PCR-amplified using gene-specific primers to contain the
         complete open reading frame (based on known gene sequences
         available from NCBI's RefSeq). Template for PCR is cDNA
         derived from either pooled cytoplasmic polyA RNA from 30
         cells lines or pooled total RNA from 10 different tissues
         (from BD Biosciences/Clontech and Washington University).
         PCR products are directionally cloned into the loxp sites
         of the pDNR-Dual vector. Library constructed by Dr.
         Narayan Bhat, Earl Bere III and Hongling Liao (Gene
         Expression Laboratory, Research Technology Program, SAIC
         Frederick, NCI-Frederick, Frederick, MD 21702). For
         information on which gene each clone represents, please
         visit our anonymous ftp site at
         ftp://image.llnl.gov/image/rearrayed_plates/IRBK.presv.dat
         a Note: this is a NIH_MGC Library."

ORIGIN
Query Match      68.7%; Score 20.6; DB 14; Length 804;
Best Local Similarity 85.2%; Pred. No. 4.6e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCGCGAGAACCGAGGTAGCACCACAT 27
    |||||
DB 99 TGCGCGTGAACCGAGGTAGCAGAGT 73
    |||||

RESULT 12
BZ773303
LOCUS
DEFINITION  BZ773303 896 bp DNA linear GSS 13-MAR-2003
          mcv67d11.g1 HFOSMID007 Homo sapiens genomic, genomic survey
          sequence.
ACCESSION  BZ773303
VERSION    BZ773303.1 GI:28946987
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 986)
AUTHORS Cook,L., Delehaunty,K., Fewell,G., Fulton,L., Magrini,V.,
Mardis,E., Miner,T., Nash,W., Williams,D. and Wilson,R.K.
TITLE Homo sapiens Fosmid End Reads
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: mcv67 row: d column: 11
Class: fosmid ends
High quality sequence start: 54
High quality sequence stop: 550.
Location/Qualifiers
1. .986
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="HFO5MID007"
/notes="Vector: pcc01fos; Site 1: Eco72i; Human whole
genome fosmid library was prepared at Washington
University Genome Sequencing Center. DNA was sheared for
blunt-ended ligation into pcc01fos inducible vector. DNA
was ordered from Coriell Cell Repository's DNA
polymorphism discovery resource."
ORIGIN
Query Match 58.7%; Score 20.6; DB 28; Length 986;
Best Local Similarity 85.2; Pred.No. 4.9e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 GCGCGAAGCAGGAGTAGGACCAATG 28
DB 175 GGGCGAAGCAGGAGTAGGACCAATG 201

RESULT 13
BF444251/c
LOCUS BP444251
DEFINITION B22053 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.
ACCESSION BF444251
VERSION BF444251.1 GI:11504343
KEYWORDS EST.
SOURCE Sus scrofa
ORGANISM Sus scrofa (pig)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 227)
AUTHORS Fahrenkrug,S.C., Smith,T.P.L., Preking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R.,
Quackenbush,J. and Keale,J.W.
TITLE Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
JOURNAL Mamm. Genome 13 (8), 475-478 (2002)
MEDLINE 22213789
PUBMED 12226715
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCCTAGTCAGGAG
Plate: 96 row: N column: 16
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1. .227
/organism="Sus scrofa"

```



Dept. of Plant Pathology  
University of Arkansas  
217 Plant Science Building, Fayetteville, AR 72701, USA  
Tel: 501 575 5191  
Fax: 501 575 7601  
Email: kkorth@comp.uark.edu  
Insert Length: 290 Std Error: 0.00  
Plate: 037 row: A column: 02  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1. .290  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF037A02IN"  
/tissue\_type="local and systemic leaves"  
/dev\_stage="mature"  
/clone\_lib="Insect herbivory"  
/notes="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

ORIGIN

Query Match 68.0%; Score 20.4; DB 10; Length 290;  
Best Local Similarity 80.0%; Pred. No. 4e+03;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCCCAATGAG 30  
|||||  
Db 196 TGGTGCAGACGACGACGACCAACAATGAG 167  
|||||

RESULT 19  
BI266129/c

LOCUS  
BI266129.1 Insect herbivory Medicago truncatula cDNA clone  
293 bp mRNA linear EST 18-JUL-2001

DEFINITION  
NF100G10IN 5', mRNA sequence.

ACCESSION  
BI266129

VERSION  
BI266129.1 GI:14869978

KEYWORDS  
EST.

SOURCE  
Medicago truncatula (barrel medic)

ORGANISM  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.

REFERENCE  
1 (bases 1 to 293)  
Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,  
Flores,H.R., Iman,J.T., Weller,J.W. and May,G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula insect herbivory library  
Unpublished (2000)

JOURNAL  
COMMENT  
Contact: Korth K  
Dept. of Plant Pathology  
University of Arkansas  
217 Plant Science Building, Fayetteville, AR 72701, USA  
Tel: 501 575 5191  
Fax: 501 575 7601  
Email: kkorth@comp.uark.edu  
Insert Length: 293 Std Error: 0.00  
Plate: 100 row: G column: 10  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1. .293  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF100G10IN"  
/tissue\_type="local and systemic leaves"  
/dev\_stage="mature"  
/clone\_lib="Insect herbivory"

FEATURES  
source

/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

## ORIGIN

Query Match 68.0%; Score 20.4; DB 12; Length 293;  
Best Local Similarity 80.0%; Pred. No. 4e+03;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCCGAGAACCCAGGAGTAGCACCACATGAG 30  
|||||  
DB 179 TGGTCGAGAGCAGCAGCAGCAACAATAAG 150  
|||||

## RESULT 20

BI267471/c  
LOCUS  
DEFINITION  
NF105E03IN1F1023 Insect herbivory Medicago truncatula cDNA clone  
VERSION  
BI267471  
KEYWORDS  
EST.  
SOURCE  
Medicago truncatula (barrel medic)

## ORIGIN

Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.

## REFERENCE

1 (bases 1 to 312)  
Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,

## AUTHORS

Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula insect herbivory library  
Unpublished (2000)

## JOURNAL

## COMMENT

Dept. of Plant Pathology  
University of Arkansas  
217 Plant Science Building, Fayetteville, AR 72701, USA  
Tel: 501 575 5191  
Fax: 501 575 7601  
Email: kkorthe@comp.uark.edu  
Insert Length: 312 Std Error: 0.00  
Plate: 105 row: E column: 03  
Seq primer: TCACACAGGAACAGCTATGAC.  
Location/Qualifiers  
1. .312

## FEATURES

## source

/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF105E03IN"  
/tissue\_type="local and systemic leaves"  
/dev\_stages="mature"

/clone\_lib="Insect herbivory"  
/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

## ORIGIN

Query Match 68.0%; Score 20.4; DB 12; Length 312;  
Best Local Similarity 80.0%; Pred. No. 4.1e-03;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

## QY

1 TGGCCGAGAACCCAGGAGTAGCACCACATGAG 30  
|||||

## DB

188 TGGTCGAGAGCAGCAGCAGCAACAATAAG 159  
|||||

## RESULT 21

BI265945/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## EST.

## Medicago truncatula (barrel medic)

## Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

## Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

## rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

## Medicago.

## 1 (bases 1 to 319)

## Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,

## Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

## Expressed Sequence Tags from the Samuel Roberts Noble Foundation

## Medicago truncatula insect herbivory library

## Unpublished (2000)

## Contact: Korth K

## Dept. of Plant Pathology

## University of Arkansas

## 217 Plant Science Building, Fayetteville, AR 72701, USA

## Tel: 501 575 5191

## Fax: 501 575 7601

## Email: kkorthe@comp.uark.edu

## Insert Length: 319 Std Error: 0.00

## Plate: 092 row: F column: 12

## Seq primer: TCACACAGGAACAGCTATGAC.

## Location/Qualifiers

## 1. .319

## /organism="Medicago truncatula"

## /mol\_type="mRNA"

## /db\_xref="taxon:3880"

## /clone="NF092F12IN"

## /tissue\_type="local and systemic leaves"

## /dev\_stages="mature"

## /clone\_lib="Insect herbivory"

## /note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## 0;

## QY

## 1

## TGGCCGAGAACCCAGGAGTAGCACCACATGAG 30

## |||||

## DB

## 197

## TGGTCGAGAGCAGCAGCAGCAACAATAAG 168

## |||||

## RESULT 22

## AW127696/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## EST.

## Medicago truncatula (barrel medic)

## Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

## Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

## rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

## Medicago.

## 1 (bases 1 to 349)

## Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P.,

## Gonzales, M.B. and Ellis, L.

## ESTs from Medicago truncatula leaves and cotyledons

## Unpublished (2000)

## On Oct 25, 1999 this sequence version replaced gi:6115600.

## Contact: Carroll P. Vance

Department of Agronomy and Plant Genetics  
University of Minnesota  
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
Tel: 612 625 5715  
Fax: 651-649-5058  
Email: vance004@maroon.tc.umn.edu  
Date: 9/15/99: Updated to the Database of Expressed Sequence Tags  
(dbEST) on 04/27/00: More information is available at  
'http://chrystie.tamu.edu/medicago'.  
Seq primer: T3.  
Location/Qualifiers  
1. 349  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/cultivar="Genotype A17"  
/db\_xref="taxon:3880"  
/clone="L270"  
/tissue\_type="leaves and cotyledons"  
/dev\_stage="mixture of cotyledons from five days old  
plants and leaves obtained from two weeks old plants"  
/lab\_host="E. coli strain SOLR"  
/clone\_lib="DSLCL"  
/note="Vector: pluescript SK +/-; Site 1: EcoRI; Site 2:  
XhoI; cDNA was prepared from polyA+ enriched RNA from the  
mixture of cotyledons of five days old plants and leaves  
of two weeks old plants. The cDNA was directionally  
ligated into the Uni-ZAP XR vector from Stratagene and  
packaged using Gigapack III Gold packaging extracts.  
Plasmids containing cDNA inserts were excised from the  
recombinant lambda-ZAP phage using EX-Assist helper phage  
and propagated in SOLR cells."

FEATURES  
source  
1. 358  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF010H11PL"  
/tissue\_type="leaf"  
/dev\_stage="trifoliolate"  
/clone\_lib="phosphate starved leaf"  
/note="Vector: Lambda Zap; At the trifoliolate stage, M.  
truncatula plants were transplanted to phosphate-free sand  
and grown for a further 30 days. During this 30 day  
period, the plants were fertilized twice weekly with 1/2  
Hoaglands solution containing only 20uM potassium  
phosphate. RNA was prepared from above ground tissues."

ORIGIN  
Query Match 68.0%; Score 20.4; DB 10; Length 358;  
Best Local Similarity 80.0%; Pred. No. 4.3e+03;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TGGCGCAGACAGGAGTAGCACCACCAATGAG 30  
|||||  
Db 203 TGGTCGAGAGCAGCAGCAGCAACAATAAG 174  
|||||

RESULT 24  
BE323556/c  
LOCUS BE323556 361 bp mRNA linear EST 21-DEC-2000  
DEFINITION NF028C04LF 5', mRNA sequence.  
ACCESSION BE323556  
VERSION BE315922  
KEYWORDS EST.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula

REFERENCE  
1 (bases 1 to 361)  
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula leaf library  
Unpublished (2000)  
On Jul 14, 2000 this sequence version replaced gi:9189699.  
Contact: May GD  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Medicago Genome Initiative accession: MGI:S:16476  
Insert Length: 772 Std Error: 0.00  
Plate: 028 row: C column: 04  
Seq primer: TCACACGAGAACAGCTATGAC.  
Location/Qualifiers  
1. 361  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF028C04LF"  
/tissue\_type="leaf"  
/dev\_stage="Pooled developmental"  
/clone\_lib="Developing leaf"  
/note="Vector: Lambda Zap; Contains a mixture of very  
young, developing, mature and senescing leaves."

ORIGIN  
Query Match 68.0%; Score 20.4; DB 10; Length 361;

Department of Agronomy and Plant Genetics  
University of Minnesota  
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
Tel: 612 625 5715  
Fax: 651-649-5058  
Email: vance004@maroon.tc.umn.edu  
Date: 9/15/99: Updated to the Database of Expressed Sequence Tags  
(dbEST) on 04/27/00: More information is available at  
'http://chrystie.tamu.edu/medicago'.  
Seq primer: T3.  
Location/Qualifiers  
1. 349  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/cultivar="Genotype A17"  
/db\_xref="taxon:3880"  
/clone="L270"  
/tissue\_type="leaves and cotyledons"  
/dev\_stage="mixture of cotyledons from five days old  
plants and leaves obtained from two weeks old plants"  
/lab\_host="E. coli strain SOLR"  
/clone\_lib="DSLCL"  
/note="Vector: pluescript SK +/-; Site 1: EcoRI; Site 2:  
XhoI; cDNA was prepared from polyA+ enriched RNA from the  
mixture of cotyledons of five days old plants and leaves  
of two weeks old plants. The cDNA was directionally  
ligated into the Uni-ZAP XR vector from Stratagene and  
packaged using Gigapack III Gold packaging extracts.  
Plasmids containing cDNA inserts were excised from the  
recombinant lambda-ZAP phage using EX-Assist helper phage  
and propagated in SOLR cells."

FEATURES  
source  
1. 358  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/db\_xref="taxon:3880"  
/clone="NF010H11PL"  
/tissue\_type="leaf"  
/dev\_stage="trifoliolate"  
/clone\_lib="phosphate starved leaf"  
/note="Vector: Lambda Zap; At the trifoliolate stage, M.  
truncatula plants were transplanted to phosphate-free sand  
and grown for a further 30 days. During this 30 day  
period, the plants were fertilized twice weekly with 1/2  
Hoaglands solution containing only 20uM potassium  
phosphate. RNA was prepared from above ground tissues."

ORIGIN  
Query Match 68.0%; Score 20.4; DB 10; Length 358;  
Best Local Similarity 80.0%; Pred. No. 4.2e+03;  
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TGGCGCAGACAGGAGTAGCACCACCAATGAG 30  
|||||  
Db 215 TGGTCGAGAGCAGCAGCAGCAACAATAAG 186  
|||||

RESULT 23  
BE323556/c  
LOCUS BE323556 358 bp mRNA linear EST 21-DEC-2000  
DEFINITION NF010H11PLF1092 Phosphate starved leaf Medicago truncatula cDNA  
clone NF010H11PL 5', mRNA sequence.  
ACCESSION BE323556  
VERSION BE323556  
KEYWORDS EST.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula

REFERENCE  
1 (bases 1 to 358)  
Lit, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula phosphate-starved leaf library  
Unpublished (2000)  
On Jul 14, 2000 this sequence version replaced gi:9197333.  
Contact: Harrison MJ  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7325  
Fax: 580 221 7380  
Email: mjharrison@noble.org  
Medicago Genome Initiative accession: MGI:S:19768  
Insert Length: 794 Std Error: 0.00  
Plate: 010 row: H column: 11







```

clone NF087A01PL 5', mRNA sequence.
ACCESSION BI263513
VERSION BI263513.1 GI:14864818
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
          Medicago truncatula
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
          rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
          Medicago.
REFERENCE 1 (bases 1 to 398)
AUTHORS Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
          Flores, H.R., Inman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula phosphate-starved leaf library
COMMENT Unpublished (2000)
          Contact: Harrison MJ
          Plant Biology Division
          The Samuel Roberts Noble Foundation
          2510 Sam Noble Parkway, Ardmore, OK 73402, USA
          Tel: 580 221 7325
          Fax: 580 221 7380
          Email: mjharrison@noble.org
          Insert Length: 398 Std Error: 0.00
          Plate: 087 row: A column: 01
          Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES
    source
        1..398
            /organism="Medicago truncatula"
            /mol_type="mRNA"
            /db_xref="taxon:3880"
            /clone="NF087A01PL"
            /tissue_type="leaf"
            /dev_stage="trifoliolate"
            /clone_lib="Phosphate starved leaf"
            /note="Vector: Lambda Zap; At the trifoliolate stage, M.
            truncatula plants were transplanted to phosphate-free sand
            and grown for a further 30 days. During this 30 day
            period, the plants were fertilized twice weekly with 1/2
            Hoaglands solution containing only 20uM potassium
            phosphate. RNA was prepared from above ground tissues."
ORIGIN
Query Match 68.0%; Score 20.4; DB 12; Length 398;
Best Local Similarity 80.0%; Pred. No. 4.4e+03;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 TGGCGCAGAACCCAGGAGTAGCCACCAATGAG 30
Db 228 TGGTCGAGAGCAGCAGCAGCAACATAG 199
Search completed: June 20, 2004, 14:15:59
Job time : 1425.9 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 09:46:51 ; Search time 33.1746 Seconds  
(without alignments)  
501.846 Million cell updates/sec

Title: US-10-624-714-18

Perfect score: 30

Sequence: 1 tggcgacagaccaggtagcaccacatgag 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length  | ID | Description         |
|------------|-------|-------------|---------|----|---------------------|
| C 1        | 30    | 100.0       | 4403765 | 3  | US-09-103-840A-2    |
| C 2        | 30    | 100.0       | 4411529 | 3  | US-09-103-840A-1    |
| C 3        | 20.6  | 68.7        | 510     | 4  | US-09-621-976-1444  |
| C 4        | 20.6  | 68.7        | 1476    | 4  | US-08-277-031B-10   |
| C 5        | 19.2  | 84.0        | 21706   | 4  | US-08-961-527-36    |
| C 6        | 18.8  | 62.7        | 1824    | 1  | US-08-305-505-1     |
| C 7        | 18.8  | 62.7        | 2907    | 4  | US-09-023-655-1053  |
| C 8        | 18.6  | 62.0        | 290     | 4  | US-09-313-294A-6733 |
| C 9        | 18.4  | 61.3        | 409     | 4  | US-09-621-976-18228 |
| C 10       | 18.4  | 61.3        | 461     | 4  | US-09-621-976-18229 |
| C 11       | 18.4  | 61.3        | 492     | 4  | US-08-621-976-3472  |
| C 12       | 18.4  | 61.3        | 538     | 4  | US-09-621-976-18226 |
| C 13       | 18.4  | 61.3        | 641     | 4  | US-09-621-976-18227 |
| C 14       | 18.4  | 61.3        | 4219    | 4  | US-09-621-976-18226 |
| C 15       | 18.2  | 60.7        | 4403765 | 3  | US-09-103-840A-2    |
| C 16       | 18.2  | 60.7        | 4411529 | 3  | US-09-103-840A-1    |
| C 17       | 18    | 60.0        | 3879    | 3  | US-08-916-352-1     |
| C 18       | 18    | 60.0        | 15998   | 4  | US-08-676-610B-24   |
| C 19       | 18    | 60.0        | 197496  | 4  | US-09-877-177A-10   |
| C 20       | 18    | 60.0        | 1830121 | 4  | US-09-557-884-1     |
| C 21       | 18    | 60.0        | 1830121 | 4  | US-09-643-990A-1    |
| C 22       | 17.8  | 59.3        | 312     | 4  | US-08-973-544-2     |
| C 23       | 17.8  | 59.3        | 5581    | 4  | US-08-973-544-1     |
| C 24       | 17.8  | 59.3        | 6911    | 1  | US-08-311-174-4     |
| C 25       | 17.6  | 58.7        | 481     | 4  | US-08-621-976-1469  |
| C 26       | 17.6  | 58.7        | 654     | 4  | US-09-184-418C-99   |
| C 27       | 17.6  | 58.7        | 1527    | 4  | US-09-543-681A-2651 |

|      |      |        |   |                      |                    |
|------|------|--------|---|----------------------|--------------------|
| 17.6 | 58.7 | 1813   | 4 | US-09-620-312D-29    | Sequence 29, Appl  |
| 17.6 | 58.7 | 9060   | 4 | US-09-184-418C-10    | Sequence 10, Appl  |
| 17.4 | 58.0 | 948    | 4 | US-09-598-401C-107   | Sequence 107, Appl |
| 17.4 | 58.0 | 1766   | 4 | US-09-598-401C-117   | Sequence 117, Appl |
| 17.4 | 58.0 | 2283   | 4 | US-09-252-991A-3298  | Sequence 3298, Ap  |
| 17.2 | 57.3 | 490    | 4 | US-09-621-976-18339  | Sequence 18339, A  |
| 17.2 | 57.3 | 1266   | 4 | US-09-543-681A-285   | Sequence 285, App  |
| 17.2 | 57.3 | 1902   | 4 | US-09-620-312D-1057  | Sequence 1057, App |
| 17.2 | 57.3 | 2571   | 4 | US-09-620-312D-302   | Sequence 302, App  |
| 17.2 | 57.3 | 2598   | 4 | US-09-612-204B-23    | Sequence 23, Appl  |
| 17.2 | 57.3 | 2830   | 2 | US-09-010-928B-1     | Sequence 1, Appli  |
| 17.2 | 57.3 | 7015   | 4 | US-09-770-315-1      | Sequence 1, Appli  |
| 17.2 | 57.3 | 80161  | 3 | US-09-036-987A-1     | Sequence 1, Appli  |
| 17.2 | 57.3 | 80161  | 3 | US-09-370-700-1      | Sequence 1, Appli  |
| 17.2 | 57.3 | 80161  | 4 | US-09-603-207-1      | Sequence 1, Appli  |
| 17   | 56.7 | 444    | 4 | US-09-621-976-18828  | Sequence 10068, A  |
| 17   | 56.7 | 445    | 4 | US-09-621-976-18828  | Sequence 18828, A  |
| 17   | 56.7 | 717    | 4 | US-09-134-001C-299   | Sequence 299, App  |
| 17   | 56.7 | 1098   | 4 | US-09-328-352-1981   | Sequence 1981, Ap  |
| 17   | 56.7 | 1509   | 4 | US-09-134-001C-761   | Sequence 761, App  |
| 17   | 56.7 | 2634   | 2 | US-08-507-166-7      | Sequence 7, Appli  |
| 17   | 56.7 | 2634   | 4 | US-09-391-340-7      | Sequence 7, Appli  |
| 17   | 56.7 | 2866   | 4 | US-09-724-864-30     | Sequence 30, Appl  |
| 17   | 56.7 | 3651   | 2 | US-08-790-374-1      | Sequence 1, Appli  |
| 17   | 56.7 | 9775   | 3 | US-08-977-171-1      | Sequence 1, Appli  |
| 16.8 | 56.0 | 668    | 3 | US-08-998-416-150    | Sequence 150, App  |
| 16.8 | 56.0 | 705    | 4 | US-09-540-236-278    | Sequence 278, App  |
| 16.8 | 56.0 | 909    | 4 | US-08-148-708-9      | Sequence 9, Appli  |
| 16.8 | 56.0 | 1179   | 4 | US-08-148-708-3      | Sequence 3, Appli  |
| 16.8 | 56.0 | 1378   | 1 | US-08-759-848-2      | Sequence 2, Appli  |
| 16.8 | 56.0 | 1378   | 4 | US-08-148-708-6      | Sequence 6, Appli  |
| 16.8 | 56.0 | 1378   | 4 | US-09-016-434-1123   | Sequence 1123, Ap  |
| 16.8 | 56.0 | 1378   | 5 | PCT-US95-09383-2     | Sequence 2, Appli  |
| 16.8 | 56.0 | 2205   | 3 | US-08-687-590-59     | Sequence 59, Appl  |
| 16.8 | 56.0 | 2574   | 2 | US-08-677-734A-8     | Sequence 8, Appli  |
| 16.8 | 56.0 | 2574   | 4 | US-09-097-053-8      | Sequence 8, Appli  |
| 16.8 | 56.0 | 3002   | 4 | US-09-369-364A-1     | Sequence 1, Appli  |
| 16.8 | 56.0 | 3577   | 4 | US-09-620-312D-457   | Sequence 457, App  |
| 16.8 | 56.0 | 6703   | 4 | US-09-596-002-7      | Sequence 7, Appli  |
| 16.8 | 56.0 | 6703   | 4 | US-08-545-528D-1     | Sequence 1, Appli  |
| 16.8 | 56.0 | 580073 | 3 | US-08-865-960-4      | Sequence 4, Appli  |
| 16.6 | 55.3 | 39     | 3 | US-09-252-991A-15319 | Sequence 15319, A  |
| 16.6 | 55.3 | 396    | 4 | US-08-346-611-1      | Sequence 1, Appli  |
| 16.6 | 55.3 | 402    | 1 | US-08-346-611-3      | Sequence 3, Appli  |
| 16.6 | 55.3 | 402    | 1 | US-08-794-494-1      | Sequence 1, Appli  |
| 16.6 | 55.3 | 402    | 2 | US-08-794-494-3      | Sequence 3, Appli  |
| 16.6 | 55.3 | 456    | 4 | US-09-252-991A-15319 | Sequence 15319, A  |
| 16.6 | 55.3 | 891    | 4 | US-09-134-001C-1636  | Sequence 1636, Ap  |
| 16.6 | 55.3 | 945    | 4 | US-09-352-991A-15367 | Sequence 15367, A  |
| 16.6 | 55.3 | 1047   | 4 | US-09-784-508-15     | Sequence 15, Appl  |
| 16.6 | 55.3 | 1074   | 4 | US-09-252-991A-4289  | Sequence 4289, Ap  |
| 16.6 | 55.3 | 1305   | 4 | US-09-242-859A-3     | Sequence 3, Appli  |
| 16.6 | 55.3 | 1305   | 4 | US-09-242-859A-7     | Sequence 7, Appli  |
| 16.6 | 55.3 | 1411   | 4 | US-09-370-838-154    | Sequence 154, App  |
| 16.6 | 55.3 | 1411   | 4 | US-09-252-991A-4561  | Sequence 4561, App |
| 16.6 | 55.3 | 1716   | 4 | US-09-252-991A-15268 | Sequence 15268, A  |
| 16.6 | 55.3 | 1848   | 4 | US-09-252-991A-15408 | Sequence 15408, A  |
| 16.6 | 55.3 | 1999   | 4 | US-09-716-129-44     | Sequence 44, Appl  |
| 16.6 | 55.3 | 2028   | 4 | US-09-489-039A-2010  | Sequence 2010, Ap  |
| 16.6 | 55.3 | 2159   | 3 | US-08-286-870A-7     | Sequence 7, Appli  |
| 16.6 | 55.3 | 2368   | 2 | US-08-520-933-1      | Sequence 1, Appli  |
| 16.6 | 55.3 | 2368   | 4 | US-09-285-930-1      | Sequence 1, Appli  |
| 16.6 | 55.3 | 2368   | 6 | 5282177-1            | Patent No. 5282177 |
| 16.6 | 55.3 | 3293   | 2 | US-08-442-809A-75    | Sequence 75, Appl  |
| 16.6 | 55.3 | 3909   | 4 | US-08-961-527-12     | Sequence 12, Appl  |
| 16.6 | 55.3 | 14113  | 3 | US-09-223-134-1      | Sequence 1, Appli  |
| 16.6 | 55.3 | 14113  | 3 | US-08-992-801-1      | Sequence 1, Appli  |
| 16.6 | 55.3 | 14113  | 3 | US-09-223-535-1      | Sequence 1, Appli  |
| 16.6 | 55.3 | 176373 | 3 | US-09-128-153-17     | Sequence 17, Appl  |
| 16.4 | 54.7 | 27     | 4 | US-09-011-336-39     | Sequence 39, Appl  |
| 16.4 | 54.7 | 60     | 3 | US-08-290-736C-37    | Sequence 37, Appl  |
| 16.4 | 54.7 | 71     | 3 | US-08-290-736C-46    | Sequence 46, Appl  |
| 16.4 | 54.7 | 71     | 3 | US-08-290-736C-47    | Sequence 47, Appl  |

|       |      |      |      |   |                     |                    |       |      |      |       |   |                      |                    |
|-------|------|------|------|---|---------------------|--------------------|-------|------|------|-------|---|----------------------|--------------------|
| C 101 | 16.4 | 54.7 | 122  | 3 | US-09-106-182-12    | Sequence 12, Appl  | C 174 | 16.4 | 54.7 | 2172  | 3 | US-09-267-439-26     | Sequence 26, Appl  |
| C 102 | 16.4 | 54.7 | 129  | 4 | US-09-589-287B-16   | Sequence 16, Appl  | C 175 | 16.4 | 54.7 | 2172  | 4 | US-09-073-138-26     | Sequence 26, Appl  |
| C 103 | 16.4 | 54.7 | 129  | 4 | US-09-268-311-10    | Sequence 10, Appl  | C 176 | 16.4 | 54.7 | 2343  | 4 | US-09-328-352-4054   | Sequence 4054, Ap  |
| C 104 | 16.4 | 54.7 | 129  | 4 | US-09-588-947A-16   | Sequence 16, Appl  | C 177 | 16.4 | 54.7 | 2425  | 4 | US-09-023-655-234    | Sequence 234, App  |
| C 105 | 16.4 | 54.7 | 129  | 4 | US-09-154-219-10    | Sequence 10, Appl  | C 178 | 16.4 | 54.7 | 2425  | 4 | US-09-254-594-5      | Sequence 5, Appl   |
| C 106 | 16.4 | 54.7 | 129  | 4 | US-09-589-286A-16   | Sequence 16, Appl  | C 179 | 16.4 | 54.7 | 3185  | 1 | US-08-117-362-2      | Sequence 2, Appl   |
| C 107 | 16.4 | 54.7 | 129  | 4 | US-09-489-039A-2612 | Sequence 2612, Ap  | C 180 | 16.4 | 54.7 | 3185  | 1 | US-08-486-924-2      | Sequence 2, Appl   |
| C 108 | 16.4 | 54.7 | 219  | 4 | US-09-621-976-12915 | Sequence 12915, A  | C 181 | 16.4 | 54.7 | 3185  | 1 | US-08-486-929A-2     | Sequence 2, Appl   |
| C 109 | 16.4 | 54.7 | 297  | 4 | US-09-621-976-13157 | Sequence 13157, A  | C 182 | 16.4 | 54.7 | 3432  | 4 | US-09-254-594-4      | Sequence 4, Appl   |
| C 110 | 16.4 | 54.7 | 299  | 4 | US-09-621-976-12708 | Sequence 12708, A  | C 183 | 16.4 | 54.7 | 3914  | 1 | US-08-117-373-11     | Sequence 11, Appl  |
| C 111 | 16.4 | 54.7 | 314  | 4 | US-09-222-575-77    | Sequence 77, Appl  | C 184 | 16.4 | 54.7 | 4104  | 1 | US-07-998-003A-94    | Sequence 94, Appl  |
| C 112 | 16.4 | 54.7 | 314  | 4 | US-09-389-681-77    | Sequence 77, Appl  | C 185 | 16.4 | 54.7 | 4104  | 1 | US-08-453-274B-94    | Sequence 94, Appl  |
| C 113 | 16.4 | 54.7 | 314  | 4 | US-09-620-405B-77   | Sequence 77, Appl  | C 186 | 16.4 | 54.7 | 4104  | 1 | US-08-453-695A-94    | Sequence 94, Appl  |
| C 114 | 16.4 | 54.7 | 314  | 4 | US-09-339-338-77    | Sequence 77, Appl  | C 187 | 16.4 | 54.7 | 4104  | 1 | US-08-268-161A-94    | Sequence 94, Appl  |
| C 115 | 16.4 | 54.7 | 314  | 4 | US-09-433-826B-77   | Sequence 77, Appl  | C 188 | 16.4 | 54.7 | 4104  | 2 | US-08-453-702A-94    | Sequence 94, Appl  |
| C 116 | 16.4 | 54.7 | 314  | 4 | US-09-604-287A-77   | Sequence 77, Appl  | C 189 | 16.4 | 54.7 | 4104  | 3 | US-09-099-639-94     | Sequence 94, Appl  |
| C 117 | 16.4 | 54.7 | 314  | 4 | US-09-285-480-77    | Sequence 77, Appl  | C 190 | 16.4 | 54.7 | 4104  | 5 | PCT-US93-12588-94    | Sequence 94, Appl  |
| C 118 | 16.4 | 54.7 | 314  | 4 | US-09-834-759-77    | Sequence 77, Appl  | C 191 | 16.4 | 54.7 | 4286  | 4 | PCT-US95-08071-94    | Sequence 94, Appl  |
| C 119 | 16.4 | 54.7 | 321  | 4 | US-09-195-106-4     | Sequence 4, Appl   | C 192 | 16.4 | 54.7 | 4286  | 4 | US-09-376-594-632    | Sequence 632, App  |
| C 120 | 16.4 | 54.7 | 353  | 4 | US-09-643-597-138   | Sequence 138, App  | C 193 | 16.4 | 54.7 | 4650  | 1 | US-07-998-003A-102   | Sequence 102, App  |
| C 121 | 16.4 | 54.7 | 353  | 4 | US-09-480-884A-138  | Sequence 138, App  | C 194 | 16.4 | 54.7 | 4650  | 1 | US-08-453-274B-102   | Sequence 102, App  |
| C 122 | 16.4 | 54.7 | 353  | 4 | US-09-542-615A-138  | Sequence 138, App  | C 195 | 16.4 | 54.7 | 4650  | 1 | US-08-453-695A-102   | Sequence 102, App  |
| C 123 | 16.4 | 54.7 | 353  | 4 | US-09-606-421B-138  | Sequence 138, App  | C 196 | 16.4 | 54.7 | 4650  | 1 | US-08-268-161A-102   | Sequence 102, App  |
| C 124 | 16.4 | 54.7 | 353  | 4 | US-09-221-107-138   | Sequence 5, Appl   | C 197 | 16.4 | 54.7 | 4650  | 2 | US-08-453-702A-102   | Sequence 102, App  |
| C 125 | 16.4 | 54.7 | 366  | 4 | US-09-329-884-5     | Sequence 5, Appl   | C 198 | 16.4 | 54.7 | 4650  | 3 | US-09-099-639-102    | Sequence 102, App  |
| C 126 | 16.4 | 54.7 | 392  | 3 | US-09-030-607-194   | Sequence 194, App  | C 199 | 16.4 | 54.7 | 4650  | 5 | PCT-US93-12588-102   | Sequence 102, App  |
| C 127 | 16.4 | 54.7 | 392  | 4 | US-09-439-313-194   | Sequence 194, App  | C 200 | 16.4 | 54.7 | 4650  | 5 | PCT-US95-08071-102   | Sequence 102, App  |
| C 128 | 16.4 | 54.7 | 392  | 4 | US-09-352-616A-194  | Sequence 194, App  | C 201 | 16.4 | 54.7 | 4680  | 1 | US-08-234-358-1      | Sequence 1, Appl   |
| C 129 | 16.4 | 54.7 | 392  | 4 | US-09-332-149A-194  | Sequence 194, App  | C 202 | 16.4 | 54.7 | 4680  | 1 | US-08-475-391-1      | Sequence 1, Appl   |
| C 130 | 16.4 | 54.7 | 392  | 4 | US-09-159-812-194   | Sequence 194, App  | C 203 | 16.4 | 54.7 | 4680  | 2 | US-08-709-609-1      | Sequence 1, Appl   |
| C 131 | 16.4 | 54.7 | 392  | 4 | US-09-636-215-194   | Sequence 194, App  | C 204 | 16.4 | 54.7 | 4680  | 5 | PCT-US95-07178-1     | Sequence 3, Appl   |
| C 132 | 16.4 | 54.7 | 392  | 4 | US-09-685-166A-194  | Sequence 194, App  | C 205 | 16.4 | 54.7 | 4771  | 3 | US-08-840-062-3      | Patent No. 549251  |
| C 133 | 16.4 | 54.7 | 392  | 4 | US-09-115-453-194   | Sequence 194, App  | C 206 | 16.4 | 54.7 | 4825  | 6 | 5459251-1            | Sequence 2, Appl   |
| C 134 | 16.4 | 54.7 | 392  | 4 | US-09-688-489-194   | Sequence 194, App  | C 207 | 16.4 | 54.7 | 4910  | 2 | US-08-331-384-2      | Sequence 2, Appl   |
| C 135 | 16.4 | 54.7 | 504  | 4 | US-09-329-884-19    | Sequence 19, Appl  | C 208 | 16.4 | 54.7 | 4910  | 2 | US-08-836-087-2      | Sequence 2, Appl   |
| C 136 | 16.4 | 54.7 | 504  | 4 | US-09-329-884-21    | Sequence 21, Appl  | C 209 | 16.4 | 54.7 | 4910  | 3 | US-09-246-320-2      | Sequence 2, Appl   |
| C 137 | 16.4 | 54.7 | 510  | 3 | US-08-432-994A-7    | Sequence 7, Appl   | C 210 | 16.4 | 54.7 | 4910  | 3 | US-09-546-738-2      | Sequence 2, Appl   |
| C 138 | 16.4 | 54.7 | 510  | 4 | US-08-250-846-7     | Sequence 7, Appl   | C 211 | 16.4 | 54.7 | 5086  | 2 | US-08-465-485A-19    | Sequence 19, Appl  |
| C 139 | 16.4 | 54.7 | 546  | 4 | US-09-134-000C-1443 | Sequence 143, Ap   | C 212 | 16.4 | 54.7 | 5086  | 2 | US-08-365-486A-14    | Sequence 14, Appl  |
| C 140 | 16.4 | 54.7 | 640  | 4 | US-09-758-282B-17   | Sequence 17, Appl  | C 213 | 16.4 | 54.7 | 5086  | 3 | US-09-080-285-19     | Sequence 19, Appl  |
| C 141 | 16.4 | 54.7 | 650  | 4 | US-09-235-451-7     | Sequence 7, Appl   | C 214 | 16.4 | 54.7 | 5086  | 4 | US-08-880-342-14     | Sequence 14, Appl  |
| C 142 | 16.4 | 54.7 | 653  | 3 | US-08-998-416-149   | Sequence 149, App  | C 215 | 16.4 | 54.7 | 5086  | 4 | US-09-724-426-19     | Sequence 19, Appl  |
| C 143 | 16.4 | 54.7 | 654  | 4 | US-09-328-352-2942  | Sequence 2942, Ap  | C 216 | 16.4 | 54.7 | 5086  | 4 | US-09-233-527-7      | Sequence 7, Appl   |
| C 144 | 16.4 | 54.7 | 657  | 4 | US-09-023-655-1261  | Sequence 1261, Ap  | C 217 | 16.4 | 54.7 | 5086  | 5 | PCT-US93-05651-4     | Sequence 2, Appl   |
| C 145 | 16.4 | 54.7 | 734  | 4 | US-09-692-081-9     | Sequence 9, Appl   | C 218 | 16.4 | 54.7 | 5086  | 5 | PCT-US93-06251-2     | Sequence 7, Appl   |
| C 146 | 16.4 | 54.7 | 810  | 4 | US-09-626-896-15    | Sequence 15, Appl  | C 219 | 16.4 | 54.7 | 5094  | 4 | US-09-234-186-7      | Patent No. 5506344 |
| C 147 | 16.4 | 54.7 | 840  | 4 | US-09-328-352-2013  | Sequence 2013, Ap  | C 220 | 16.4 | 54.7 | 5104  | 6 | 5506344-1            | Sequence 1, Appl   |
| C 148 | 16.4 | 54.7 | 856  | 4 | US-09-328-352-591   | Sequence 591, App  | C 221 | 16.4 | 54.7 | 7214  | 4 | US-09-438-268-1      | Sequence 1, Appl   |
| C 149 | 16.4 | 54.7 | 1053 | 4 | US-09-220-132-162   | Sequence 162, App  | C 222 | 16.4 | 54.7 | 7379  | 4 | US-09-341-587-5      | Sequence 5, Appl   |
| C 150 | 16.4 | 54.7 | 1128 | 3 | US-09-107-532A-16   | Sequence 16, Appl  | C 223 | 16.4 | 54.7 | 7557  | 4 | US-09-770-315-3      | Sequence 3, Appl   |
| C 151 | 16.4 | 54.7 | 1128 | 4 | US-08-795-473B-1    | Sequence 1, Appl   | C 224 | 16.4 | 54.7 | 8067  | 1 | US-08-117-362-8      | Sequence 8, Appl   |
| C 152 | 16.4 | 54.7 | 1128 | 4 | US-09-439-856-1     | Sequence 1, Appl   | C 225 | 16.4 | 54.7 | 8067  | 1 | US-08-486-924-8      | Sequence 8, Appl   |
| C 153 | 16.4 | 54.7 | 1128 | 4 | US-09-686-344-30    | Sequence 30, Appl  | C 226 | 16.4 | 54.7 | 8151  | 4 | US-08-486-929A-8     | Sequence 2, Appl   |
| C 154 | 16.4 | 54.7 | 1128 | 4 | US-09-626-896-15    | Sequence 15, Appl  | C 227 | 16.4 | 54.7 | 8179  | 4 | US-09-438-268-2      | Sequence 5, Appl   |
| C 155 | 16.4 | 54.7 | 1145 | 6 | 5510472-1           | Patent No. 5510472 | C 228 | 16.4 | 54.7 | 8698  | 4 | US-09-438-268-5      | Sequence 2, Appl   |
| C 156 | 16.4 | 54.7 | 1155 | 4 | US-09-970-516-1     | Sequence 1, Appl   | C 229 | 16.4 | 54.7 | 8698  | 4 | US-09-770-315-2      | Sequence 1, Appl   |
| C 157 | 16.4 | 54.7 | 1192 | 4 | US-09-692-081-1     | Sequence 1, Appl   | C 230 | 16.4 | 54.7 | 15462 | 3 | US-08-976-063E-1     | Sequence 1, Appl   |
| C 158 | 16.4 | 54.7 | 1215 | 4 | US-09-489-039A-1714 | Sequence 1714, Ap  | C 231 | 16.4 | 54.7 | 32679 | 4 | US-09-073-492-1      | Sequence 1, Appl   |
| C 159 | 16.4 | 54.7 | 1244 | 4 | US-08-328-352-3384  | Sequence 3384, Ap  | C 232 | 16.2 | 54.0 | 154   | 4 | US-09-590-759-9      | Sequence 9, Appl   |
| C 160 | 16.4 | 54.7 | 1446 | 4 | US-08-976-063B-27   | Sequence 27, Appl  | C 233 | 16.2 | 54.0 | 154   | 4 | US-09-590-759-10     | Sequence 10, Appl  |
| C 161 | 16.4 | 54.7 | 1533 | 4 | US-09-205-258-90    | Sequence 90, Appl  | C 234 | 16.2 | 54.0 | 262   | 4 | US-09-313-294A-711   | Sequence 711, App  |
| C 162 | 16.4 | 54.7 | 1609 | 4 | US-09-376-594-850   | Sequence 850, App  | C 235 | 16.2 | 54.0 | 287   | 4 | US-08-634-137-23     | Sequence 23, Appl  |
| C 163 | 16.4 | 54.7 | 1708 | 4 | US-09-919-172-30    | Sequence 30, Appl  | C 236 | 16.2 | 54.0 | 431   | 3 | US-08-950-823-111    | Sequence 111, App  |
| C 164 | 16.4 | 54.7 | 1709 | 2 | US-09-010-398-2     | Sequence 2, Appl   | C 237 | 16.2 | 54.0 | 431   | 3 | US-09-477-135A-111   | Sequence 111, App  |
| C 165 | 16.4 | 54.7 | 1709 | 3 | US-09-366-260-2     | Sequence 2, Appl   | C 238 | 16.2 | 54.0 | 457   | 4 | US-09-621-976-11822  | Sequence 11822, A  |
| C 166 | 16.4 | 54.7 | 1734 | 4 | US-09-153-804-11    | Sequence 11, Appl  | C 239 | 16.2 | 54.0 | 460   | 4 | US-09-621-976-11822  | Sequence 11822, A  |
| C 167 | 16.4 | 54.7 | 1874 | 4 | US-09-489-039A-1599 | Sequence 1599, Ap  | C 240 | 16.2 | 54.0 | 492   | 4 | US-09-252-991A-8970  | Sequence 8970, Ap  |
| C 168 | 16.4 | 54.7 | 2115 | 4 | US-08-388-852B-1    | Sequence 891, App  | C 241 | 16.2 | 54.0 | 498   | 3 | US-09-085-199B-36    | Sequence 36, Appl  |
| C 169 | 16.4 | 54.7 | 2130 | 3 | US-09-056-105-1     | Sequence 1, Appl   | C 242 | 16.2 | 54.0 | 522   | 4 | US-09-252-991A-8750  | Sequence 8750, Ap  |
| C 170 | 16.4 | 54.7 | 2131 | 4 | US-08-234-784B-91   | Sequence 91, Appl  | C 243 | 16.2 | 54.0 | 522   | 4 | US-09-252-991A-12663 | Sequence 12663, A  |
| C 171 | 16.4 | 54.7 | 2172 | 2 | US-08-417-174-26    | Sequence 26, Appl  | C 244 | 16.2 | 54.0 | 544   | 4 | US-09-621-976-2695   | Sequence 2695, Ap  |
| C 172 | 16.4 | 54.7 | 2172 | 2 | US-08-231-565A-26   | Sequence 26, Appl  | C 245 | 16.2 | 54.0 | 636   | 4 | US-09-252-991A-9394  | Sequence 9394, Ap  |
| C 173 | 16.4 | 54.7 | 2172 | 2 | US-09-007-961-26    | Sequence 26, Appl  | C 246 | 16.2 | 54.0 | 771   | 2 | US-08-208-005C-1     | Sequence 1, Appl   |

247 16.2 54.0 771 2 US-09-038-597A-1 Sequence 1, Appli  
248 16.2 54.0 771 2 US-08-431-117A-1 Sequence 1, Appli  
C 249 16.2 54.0 780 3 US-08-855-825-11 Sequence 11, Appl  
C 250 16.2 54.0 780 3 US-08-855-825-13 Sequence 13, Appl  
C 251 16.2 54.0 780 3 US-09-252-991A-9118 Sequence 9118, Ap  
C 252 16.2 54.0 792 1 US-08-578-709-10 Sequence 10, Appl  
C 253 16.2 54.0 984 4 US-09-221-017B-708 Sequence 708, App  
C 254 16.2 54.0 1003 4 US-09-634-137-31 Sequence 31, Appl  
C 255 16.2 54.0 1089 4 US-09-107-532A-3439 Sequence 3439, Ap  
C 256 16.2 54.0 1122 4 US-09-252-991A-12729 Sequence 12729, A  
C 257 16.2 54.0 1182 4 US-09-252-991A-12886 Sequence 12886, A  
C 258 16.2 54.0 1182 4 US-09-107-532A-58 Sequence 58, Appl  
C 259 16.2 54.0 1316 4 US-09-461-325-91 Sequence 91, Appl  
C 260 16.2 54.0 1325 4 US-10-012-542-91 Sequence 91, Appl  
C 261 16.2 54.0 1365 4 US-09-023-655-1343 Sequence 1343, Ap  
C 262 16.2 54.0 1365 4 US-09-328-352-9222 Sequence 922, App  
C 263 16.2 54.0 1431 4 US-09-252-991A-10507 Sequence 10507, A  
C 264 16.2 54.0 1547 4 US-09-620-312D-966 Sequence 966, App  
C 265 16.2 54.0 1566 4 US-09-252-991A-8859 Sequence 8859, Ap  
C 266 16.2 54.0 1619 4 US-09-522-714-11 Sequence 11, Appl  
C 267 16.2 54.0 1674 4 US-09-489-039A-3633 Sequence 3633, Ap  
C 268 16.2 54.0 1866 4 US-09-252-991A-9439 Sequence 9439, Ap  
C 269 16.2 54.0 1897 4 US-09-556-870A-1 Sequence 1, Appli  
C 270 16.2 54.0 1845 4 US-09-540-236-938 Sequence 938, App  
C 271 16.2 54.0 1878 4 US-09-134-000C-667 Sequence 667, App  
C 272 16.2 54.0 1977 1 US-08-578-709-14 Sequence 14, Appl  
C 273 16.2 54.0 2028 4 US-09-252-991A-12500 Sequence 12500, A  
C 274 16.2 54.0 2040 4 US-09-252-991A-12977 Sequence 12977, A  
C 275 16.2 54.0 2079 4 US-09-489-039A-3593 Sequence 3593, Ap  
C 276 16.2 54.0 2088 4 US-09-458-481B-3 Sequence 3, Appli  
C 277 16.2 54.0 2190 1 US-09-625-188-19 Sequence 19, Appl  
C 278 16.2 54.0 2301 1 US-08-306-691B-23 Sequence 23, Appl  
C 279 16.2 54.0 2301 5 PCT-US93-06251-78 Sequence 3, Appl  
C 280 16.2 54.0 2301 5 PCT-US93-06251-78 Sequence 78, Appl  
C 281 16.2 54.0 2697 4 US-09-620-312D-401 Sequence 401, App  
C 282 16.2 54.0 2858 4 US-09-688-078-5 Sequence 5, Appli  
C 283 16.2 54.0 3098 3 US-09-232-200-58 Sequence 58, Appl  
C 284 16.2 54.0 3098 4 US-09-232-197-58 Sequence 58, Appl  
C 285 16.2 54.0 3098 4 US-09-232-201-58 Sequence 58, Appl  
C 286 16.2 54.0 3136 4 US-09-443-184-43 Sequence 43, Appl  
C 287 16.2 54.0 3136 4 US-09-232-195-58 Sequence 58, Appl  
C 288 16.2 54.0 3757 4 US-09-023-655-1502 Sequence 1502, Ap  
C 289 16.2 54.0 3999 4 US-09-252-991A-10309 Sequence 10309, A  
C 290 16.2 54.0 4095 4 US-08-308-881-5 Sequence 5, Appli  
C 291 16.2 54.0 4171 1 US-09-058-263-5 Sequence 5, Appli  
C 292 16.2 54.0 4171 2 US-09-059-099-5 Sequence 5, Appli  
C 293 16.2 54.0 4171 3 US-09-058-264-5 Sequence 5, Appli  
C 294 16.2 54.0 4171 4 US-09-455-962-5 Sequence 5, Appli  
C 295 16.2 54.0 4171 5 PCT-US93-06530-5 Sequence 5, Appli  
C 296 16.2 54.0 6426 4 US-09-976-594-136 Patent No. 5223423  
C 297 16.2 54.0 9633 6 5223423-1 Sequence 136, App  
C 298 16.2 54.0 9633 6 5223423-1 Sequence 136, App  
C 299 16.2 54.0 148567 4 US-09-801-876B-3 Sequence 3, Appli  
C 300 16.2 54.0 148567 4 US-10-254-869-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24

US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24

Query Match 100.0%; Score 30; DB 3; Length 4403765;  
Best Local Similarity 100.0%; Pred. No. 0.0061;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30  
Db 24986 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 24987

RESULT 2  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24

Query Match 100.0%; Score 30; DB 3; Length 4411529;  
Best Local Similarity 100.0%; Pred. No. 0.0061;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 30  
Db 25004 TGGCGCAGAACCCAGGAGTAGCACCACATGAG 24975

RESULT 3  
US-09-621-976-1444/c  
; Sequence 1444, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 1444  
; LENGTH: 510  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS

LOCATION: 58..363  
NAME/KEY: sig\_peptide  
LOCATION: 58..111  
OTHER INFORMATION: Von Heijne matrix  
OTHER INFORMATION: score 7.6999980926514  
OTHER INFORMATION: seq LLFLALLTGLLL/LV  
US-09-621-976-1444

Query Match 68.7%; Score 20.6; DB 4; Length 510;  
Best Local Similarity 85.2%; Pred. No. 8.8;  
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
DB 125 TGGCGCTGAACCCAGGAGTAGCAGAGCT 99

RESULT 4  
US-08-277-031B-10/c  
Sequence 10, Application US/08277031B  
Patent No. 6620593  
GENERAL INFORMATION:  
APPLICANT: Hayashi, Koji  
APPLICANT: Sakaki, Toshiyuki  
APPLICANT: Yabusaki, Yoshiyasu  
APPLICANT: Komai, Koichiro  
APPLICANT: Kaneko, Hideo  
APPLICANT: Nakatsuka, Iwao  
TITLE OF INVENTION: METHOD FOR SAFETY EVALUATION OF  
TITLE OF INVENTION: CHEMICAL COMPOUND USING RECOMBINANT YEAST EXPRESSING  
TITLE OF INVENTION: HUMAN CYTOCHROME P450  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5-inch, 1.44MB  
COMPUTER: IBM PC  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/277,031B  
FILING DATE: 19-JULY-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-201120/1993  
APPLICATION NUMBER: JP-180246/1993  
APPLICATION NUMBER: JP-208279/1993  
FILING DATE: 20-07-1993  
FILING DATE: 21-07-1993  
FILING DATE: 30-07-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Raymond C. Stewart  
REGISTRATION NUMBER: 21,066  
REFERENCE/DOCKET NUMBER: 20-3530P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
TELEX:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1476  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-277-031B-10

Query Match 68.7%; Score 20.6; DB 4; Length 1476;  
Best Local Similarity 85.2%; Pred. No. 11;

Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACAT 27  
DB 68 TGGCGCTGAACCCAGGAGTAGCAGAGCT 42

RESULT 5  
US-08-961-527-36  
Sequence 36, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21706 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-36

Query Match 64.0%; Score 19.2; DB 4; Length 21706;  
Best Local Similarity 87.5%; Pred. No. 72;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCAGAACCCAGGAGTAGCACCACATG 28  
DB 6961 GAAGAACCCAGGAGTAGCTCCATG 6984

RESULT 6  
US-08-305-505-1/c  
Sequence 1, Application US/08305505  
Patent No. 5668001  
GENERAL INFORMATION:  
APPLICANT: Mizutoko, Henry M.  
TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA  
TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED  
TITLE OF INVENTION: STABILITY  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 411 East Wisconsin Avenue  
CITY: Milwaukee  
STATE: Wisconsin



```
; SEQ ID NO 18228
; LENGTH: 409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 55
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-18228

Query Match          61.3%; Score 18.4; DB 4; Length 409;
Best Local Similarity 78.6%; Pred. No. 71;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 237 TGGCCAGCAGCAGGACTGGCGGCAATG 264

RESULT 10
US-09-621-976-18229
; Sequence 18229, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18229
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18229

Query Match          61.3%; Score 18.4; DB 4; Length 461;
Best Local Similarity 78.6%; Pred. No. 72;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 238 TGGCCAGCAGCAGGACTGGCGGCAATG 265

RESULT 11
US-09-621-976-3472
; Sequence 3472, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 3472
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 156..347
US-09-621-976-3472

Query Match          61.3%; Score 18.4; DB 4; Length 492;
Best Local Similarity 78.6%; Pred. No. 73;
```

```
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 289 TGGCCAGCAGCAGGACTGGCGGCAATG 316

RESULT 12
US-09-621-976-18226
; Sequence 18226, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18226
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18226

Query Match          61.3%; Score 18.4; DB 4; Length 538;
Best Local Similarity 78.6%; Pred. No. 75;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 336 TGGCCAGCAGCAGGACTGGCGGCAATG 363

RESULT 13
US-09-621-976-18227
; Sequence 18227, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18227
; LENGTH: 641
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18227

Query Match          61.3%; Score 18.4; DB 4; Length 641;
Best Local Similarity 78.6%; Pred. No. 77;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCACATG 28
    |||||
Db 297 TGGCCAGCAGCAGGACTGGCGGCAATG 324

RESULT 14
US-09-621-976-594-752
; Sequence 752, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
```



```
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 752
; LENGTH: 4219
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1330149.46
; NAME/KEY: unsure
; LOCATION: 1665-1703, 3787, 3883
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-752

Query Match      61.3%; Score 18.4; DB 4; Length 4219;
Best Local Similarity 78.6%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      2 GCGCAGACAGGAGTAGCACCATGA 29
DB      134 GGAGCAGATCCAGGAGAGCTCATATGA 161

RESULT 15
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      60.7%; Score 18.2; DB 3; Length 4403765;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GCGCAGACAGGAGTAGCACC 24
DB      3660571 GGCGAAGACAGGAGTGTACC 3660593

RESULT 16
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
```

```
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      60.7%; Score 18.2; DB 3; Length 4411529;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GGCGCAGACAGGAGTAGCACC 24
DB      3665785 GGCGAAGACAGGAGTGTACC 3665807

RESULT 17
US-08-916-352-1
; Sequence 1, Application US/08916352
; Patent No. 6166191
; GENERAL INFORMATION:
; APPLICANT: CHIRON CORPORATION
; TITLE OF INVENTION: HUMAN POLYHOMEOTIC 1 (hph1) ACTS AS A
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: 4560 HORTON STREET
; CITY: EMERYVILLE
; STATE: CA
; COUNTRY: USA
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/916,352
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: POTTER, JANE
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 1355.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-923-2707
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 1
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-916-352-1

Query Match      60.0%; Score 18; DB 3; Length 3879;
Best Local Similarity 80.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 TGGCGCAGAACCCAGGAGTAGCACC 26
DB      1268 TGGCGCAGCAGCAGCAGCAGCAACAA 1293

RESULT 18
```

US-09-676-610B-24/c  
; Sequence 24, Application US/09676610B  
; Patent No. 644465  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; APPLICANT: Jacqueline Wyatt  
; APPLICANT: Susan M. Priester  
; TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION  
; FILE REFERENCE: KTS-0138  
; CURRENT APPLICATION NUMBER: US/09/676, 610B  
; CURRENT FILING DATE: 2000-09-29  
; NUMBER OF SEQ ID NOS: 182  
; SEQ ID NO 24  
; LENGTH: 169998  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: exon  
; LOCATION: (1208)...(1472)  
; NAME/KEY: intron  
; LOCATION: (1473)...(124390)  
; NAME/KEY: exon  
; LOCATION: (124391)...(124544)  
; NAME/KEY: intron  
; LOCATION: (124545)...(125409)  
; NAME/KEY: exon  
; LOCATION: (125410)...(125595)  
; NAME/KEY: intron  
; LOCATION: (125596)...(128711)  
; NAME/KEY: exon  
; LOCATION: (128712)...(128848)  
; NAME/KEY: intron  
; LOCATION: (128849)...(133400)  
; NAME/KEY: exon  
; LOCATION: (133401)...(133469)  
; NAME/KEY: intron  
; LOCATION: (133470)...(134652)  
; NAME/KEY: exon  
; LOCATION: (134653)...(134773)  
; NAME/KEY: intron  
; LOCATION: (134774)...(136116)  
; NAME/KEY: exon  
; LOCATION: (136117)...(136261)  
; NAME/KEY: intron  
; LOCATION: (136262)...(137936)  
; NAME/KEY: exon  
; LOCATION: (137937)...(138053)  
; NAME/KEY: intron  
; LOCATION: (138054)...(138637)  
; NAME/KEY: exon  
; LOCATION: (138638)...(138766)  
; NAME/KEY: intron  
; LOCATION: (138767)...(138864)  
; NAME/KEY: exon  
; LOCATION: (138865)...(138940)  
; NAME/KEY: intron  
; LOCATION: (138941)...(139765)  
; NAME/KEY: exon  
; LOCATION: (139766)...(139860)  
; NAME/KEY: intron  
; LOCATION: (139861)...(142245)  
; NAME/KEY: exon  
; LOCATION: (142246)...(142445)  
; NAME/KEY: intron  
; LOCATION: (142446)...(143605)  
; NAME/KEY: exon  
; LOCATION: (143606)...(143738)  
; NAME/KEY: intron  
; LOCATION: (143739)...(145838)  
; NAME/KEY: exon  
; LOCATION: (145839)...(145931)  
; NAME/KEY: intron  
; LOCATION: (145932)...(147385)

; NAME/KEY: exon  
; LOCATION: (147386)...(147544)  
; NAME/KEY: intron  
; LOCATION: (147545)...(153274)  
; NAME/KEY: exon  
; LOCATION: (153275)...(153321)  
; NAME/KEY: intron  
; LOCATION: (153322)...(155088)  
; NAME/KEY: exon  
; LOCATION: (155089)...(155231)  
; NAME/KEY: intron  
; LOCATION: (155232)...(156025)  
; NAME/KEY: exon  
; LOCATION: (156026)...(156151)  
; NAME/KEY: intron  
; LOCATION: (156152)...(156826)  
; NAME/KEY: exon  
; LOCATION: (156827)...(156928)  
; NAME/KEY: intron  
; LOCATION: (156929)...(163399)  
; NAME/KEY: exon  
; LOCATION: (163400)...(163586)  
US-09-676-610B-24

Query Match 60.0%; Score 18; DB 4; Length 169998;  
Best Local Similarity 80.8%; Pred. No. 3.3e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GCAGACCGAGGATGACCAATGAG 30  
|||||  
Db 107252 GCAGAGCTGGATGCCCACTGAG 107227

RESULT 19  
US-09-877-177A-10/c  
; Sequence 10, Application US/09877177A  
; Patent No. 6582919  
; GENERAL INFORMATION:  
; APPLICANT: K. Danenberg  
; TITLE OF INVENTION: Method of determining Epidermal Growth  
; TITLE OF INVENTION: Factor Receptor and HER2-Neu Gene Expression  
; TITLE OF INVENTION: and Correlation of Levels Thereof With Survival  
; FILE REFERENCE: 11220/120  
; CURRENT APPLICATION NUMBER: US/09/877,177A  
; CURRENT FILING DATE: 2001-06-11  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 197496  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-877-177A-10

Query Match 60.0%; Score 18; DB 4; Length 197496;  
Best Local Similarity 80.8%; Pred. No. 3.4e+02;  
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GCAGACCGAGGATGACCAATGAG 30  
|||||  
Db 115252 GCAGAGCTGGATGCCCACTGAG 115227

RESULT 20  
US-09-557-884-1  
; Sequence 1, Application US/09557884  
; Patent No. 6506591  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:

```
;
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 60.0%; Score 18; DB 4; Length 1830121;
Best Local Similarity 80.8%; Pred. No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGGCGCAGACCAAGAGTAGACCA 26
Db 1027675 TGCTGCAGCAGCAGTAGTAGCAGCA 1027700

RESULT 21
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
;
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
;
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 60.0%; Score 18; DB 4; Length 1830121;
Best Local Similarity 80.8%; Pred. No. 2.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 TGGCGCAGACCAAGAGTAGACCA 26
Db 1027675 TGCTGCAGCAGCAGTAGTAGCAGCA 1027700

RESULT 22
US-08-973-544-2
; Sequence 2, Application US/08973544
; Patent No. 6338950
; GENERAL INFORMATION:
; APPLICANT: WEISS, Elisabeth
; TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAI DO, MARCELSTEIN, MURRAY & ORAM LLP
; STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,544
; FILING DATE: 18-DEC-1997
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT EP 96/02663
; FILING DATE: 20-JUN-1996
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95109511.6
; FILING DATE: 20-JUN-1995
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95112201.9
; FILING DATE: 03-AUG-1995
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitts, Monica Chin
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P8341-7073
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-5000
; TELEFAX: (202) 638-4810
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 base pairs
```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..312
US-08-973-544-2

Query Match          59.3%; Score 17.8; DB 4; Length 312;
Best Local Similarity 75.9%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY  2 GGCGCAGAACCCAGGAGTAGCACCACCAATGAG 30
    |||||
Db   239 GGCGCAGACACAGAGAGGCGACCAAGGAG 267

RESULT 23
US-08-973-544-1
; Sequence 1, Application US/08973544
; Patent No. 6338950
; GENERAL INFORMATION:
; APPLICANT: WEISS, Elisabeth
; TITLE OF INVENTION: NEW IMMUNOREGULATORY PROTEIN LST-1
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIKAI, DO, MARCELSTEIN, MURRAY & CRAM LLP
; STREET: 655 Fifteenth St., NW, Suite 300, G St. Lobby
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT EP 96/02663
; FILING DATE: 20-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95109511.6
; FILING DATE: 20-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95112201.9
; FILING DATE: 03-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitts, Monica Chin
; REGISTRATION NUMBER: 36,105
; REFERENCE/DOCKET NUMBER: P8341-7073
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-5000
; TELEFAX: (202) 638-4810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 48..162
; NAME/KEY: exon
; LOCATION: 544..652
; NAME/KEY: exon

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..312
US-08-973-544-2

Query Match          59.3%; Score 17.8; DB 4; Length 312;
Best Local Similarity 75.9%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY  2 GGCGCAGAACCCAGGAGTAGCACCACCAATGAG 30
    |||||
Db   239 GGCGCAGACACAGAGAGGCGACCAAGGAG 267

RESULT 24
US-08-311-174-4/c
; Sequence 4, Application US/083111174
; Patent No. 5556776
; GENERAL INFORMATION:
; APPLICANT: TSUCHIYA, MAKOTO
; APPLICANT: MIWA, KIYOSHI
; TITLE OF INVENTION: SUCRASE GENE DERIVED FROM CORYNEFORM
; TITLE OF INVENTION: BACTERIA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,174
; FILING DATE: 23-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 046836/1992
; FILING DATE: 04-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5556776man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-699-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6911 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 59.3%; Score 17.8; DB 1; Length 6911;
; Best Local Similarity 75.9%; Pred. No. 2.2e+02;
; Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY  2 GGCGCAGAACCCAGGAGTAGCACCACCAATGAG 30
    |||||
Db   2428 GGCGCAGACACAGAGAGGCGACCAAGGAG 2456
```

DB 2963 GGAGCAGAACCGAGGTGTGTCATCACTGAG 2935  
|||||

## RESULT 25

US-09-621-976-1469  
; Sequence 1469, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 1469  
; LENGTH: 481  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 100..441  
; NAME/KEY: sig.peptide  
; LOCATION: 100..303  
; OTHER INFORMATION: Von Heijne matrix  
; OTHER INFORMATION: score 7.5  
; OTHER INFORMATION: seq VLMFSQVISICWA/AM  
US-09-621-976-1469

Query Match 58.7%; Score 17.6; DB 4; Length 481;

Best Local Similarity 83.3%; Pred. No. 1.6e+02; Mismatches 4; Indels 0; Gaps 0;

QY 7 AGAACGAGGAGTAGCACCACATGAG 30  
|||||

DB 396 AGTACCGGAGGAGCACCACATGAG 419  
|||||

## RESULT 26

US-09-184-418C-99  
; Sequence 99, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 99  
; LENGTH: 654  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: isolate=94CY017.41; gene=nef  
US-09-184-418C-99

Query Match 58.7%; Score 17.6; DB 4; Length 654;

Best Local Similarity 83.3%; Pred. No. 1.17e+02; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCAGACCGAGGAGTAGCACCACATG 28  
|||||

DB 106 GCAGACCGAGGAGTAGGAGCAGTG 129  
|||||

## RESULT 27

Query Match 58.7%; Score 17.6; DB 4; Length 1813;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02; Mismatches 4; Indels 0; Gaps 0;

US-09-543-681A-2651/c  
; Sequence 2651, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:

; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 2651  
; LENGTH: 1527  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-2651

Query Match 58.7%; Score 17.6; DB 4; Length 1527;

Best Local Similarity 83.3%; Pred. No. 2e+02; Mismatches 4; Indels 0; Gaps 0;

QY 4 CGCAGAACCGAGGAGTAGCACCACAT 27  
|||||

DB 828 CCCAGTCCAGGAATAGCACCACAT 805  
|||||

## RESULT 28

US-09-620-312D-29  
; Sequence 29, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yuning  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radcoje T.  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 29  
; LENGTH: 1813  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (221)..(1189)  
US-09-620-312D-29

Query Match 58.7%; Score 17.6; DB 4; Length 1813;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02; Mismatches 4; Indels 0; Gaps 0;

Db 767 GAGCAGCAGCAGTAGGAGCAGCAATGA 741

Search completed: June 20, 2004, 11:47:23  
Job time : 50.2996 secsQY 5 GCAGAACCCAGGAGTAGCACCACCAATG 28  
Db 258 GCAGTAGCAGCAGCAGCACCACCAATG 281

RESULT 29  
US-09-184-418C-10  
; Sequence 10, Application US/09184418C  
; Patent No. 6492110  
; GENERAL INFORMATION:  
; APPLICANT: Hahn, Beatrice  
; APPLICANT: Gao, Feng  
; APPLICANT: Shaw, George  
; TITLE OF INVENTION: CLONES AND SEQUENCES FOR NON-SUBTYPE B ISOLATES OF HUMAN  
; TITLE OF INVENTION: IMMUNODEFICIENCY VIRUS TYPE 1  
; FILE REFERENCE: D6287  
; CURRENT APPLICATION NUMBER: US/09/184,418C  
; CURRENT FILING DATE: 1999-11-02  
; NUMBER OF SEQ ID NOS: 112  
; SEQ ID NO 10  
; LENGTH: 9060  
; TYPE: DNA  
; ORGANISM: Human immunodeficiency virus type 1  
; FEATURE:  
; OTHER INFORMATION: Isolates=94CY017.41; 159.1649:"gag";  
; OTHER INFORMATION: 1442.4453:"pol"; 4398.4976:"vif"; 4916.5206:"vpr";  
; OTHER INFORMATION: 5187.7841:"tat"; 5326.8046:"rev";  
; OTHER INFORMATION: 5428.5673:"vpu"; 5591.8188:"env"; 8190.8843:"nsf"  
US-09-184-418C-10

Query Match 58.7%; Score 17.6; DB 4; Length 9060;  
Best Local Similarity 83.3%; Pred.No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCAGAACCCAGGAGTAGCACCACCAATG 28  
Db 8295 GCAGCAGCAGGAGTAGGAGCAGATG 8318

RESULT 30  
US-09-598-401C-107/c  
; Sequence 107, Application US/09598401C  
; Patent No. 6596925  
; GENERAL INFORMATION:  
; APPLICANT: Perera, J. Ranjan  
; APPLICANT: Eagleton, Claire  
; APPLICANT: Rice, Stephen J.  
; TITLE OF INVENTION: Compositions and Methods for the  
; TITLE OF INVENTION: Modification of Gene Expression  
; FILE REFERENCE: 11000.1036c2  
; CURRENT APPLICATION NUMBER: US/09/598,401C  
; CURRENT FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: U.S. No. 6596925 60/146,591  
; PRIOR FILING DATE: 1999-07-30  
; PRIOR APPLICATION NUMBER: PCT/NZ00/00018  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: U.S. No. 6596925 09/276,599  
; PRIOR FILING DATE: 1999-03-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 107  
; LENGTH: 948  
; TYPE: DNA  
; ORGANISM: Eucalyptus grandis  
US-09-598-401C-107

Query Match 58.0%; Score 17.4; DB 4; Length 948;  
Best Local Similarity 77.8%; Pred.No. 2.2e+02;  
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 GCAGAACCCAGGAGTAGCACCACCAATGA 29

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 10:18:27 ; Search time 159.683 Seconds  
(without alignments)  
809.940 Million cell updates/sec

Title: US-10-624-714-18  
Perfect score: 30  
Sequence: 1 tggccagaacaggtagaccacatgag 30

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 300 summaries

Database : Published Applications NA.\*

- 1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 10: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 14: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 15: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 16: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 17: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 18: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 19: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 30    | 100.0       | 1584   | 9  | US-09-712-163-10    |
| 2          | 20.6  | 68.7        | 176    | 15 | US-10-029-386-25513 |
| 3          | 20.6  | 68.7        | 501    | 15 | US-10-029-386-11813 |
| 4          | 20.6  | 68.7        | 3045   | 9  | US-09-954-456-218   |
| 5          | 20.6  | 68.7        | 3045   | 9  | US-09-880-107-2319  |
| 6          | 20.6  | 68.7        | 3045   | 13 | US-10-342-887-400   |
| 7          | 20.6  | 68.7        | 3045   | 13 | US-10-172-118-400   |
| 8          | 20.6  | 68.7        | 3050   | 13 | US-10-342-887-525   |
| 9          | 20.6  | 68.7        | 3050   | 13 | US-10-172-118-525   |
| 10         | 20.6  | 68.7        | 42547  | 15 | US-10-268-822-12    |
| 11         | 19.6  | 65.3        | 629    | 13 | US-10-027-632-22664 |
| 12         | 19.6  | 65.3        | 629    | 16 | US-10-027-632-22664 |
| 13         | 19.6  | 65.3        | 99934  | 16 | US-10-085-117-73    |
| 14         | 19.4  | 64.7        | 360    | 16 | US-10-062-674-960   |

|    |      |      |         |    |                      |                    |
|----|------|------|---------|----|----------------------|--------------------|
| 15 | 19.4 | 64.7 | 537     | 9  | US-09-864-761-7559   | Sequence 7559, App |
| 16 | 19.4 | 64.7 | 1260    | 16 | US-10-371-264-15     | Sequence 15, Appl  |
| 17 | 19.4 | 64.7 | 1260    | 16 | US-10-371-099-15     | Sequence 15, Appl  |
| 18 | 19.4 | 64.7 | 1260    | 16 | US-10-371-122-15     | Sequence 15, Appl  |
| 19 | 19.4 | 64.7 | 1260    | 16 | US-10-373-567-15     | Sequence 15, Appl  |
| 20 | 19.4 | 64.7 | 1260    | 17 | US-10-628-088-15     | Sequence 15, Appl  |
| 21 | 19.4 | 64.7 | 1779    | 16 | US-10-369-493-25221  | Sequence 25221, A  |
| 22 | 19.4 | 64.0 | 1342    | 13 | US-10-425-114-15923  | Sequence 15923, A  |
| 23 | 19.2 | 64.0 | 1950    | 16 | US-10-369-493-46459  | Sequence 46459, A  |
| 24 | 19.2 | 64.0 | 21706   | 13 | US-10-158-844-36     | Sequence 36, Appl  |
| 25 | 19   | 63.3 | 393     | 10 | US-09-918-998-5024   | Sequence 5024, Ap  |
| 26 | 19   | 63.3 | 637     | 9  | US-09-563-817-317    | Sequence 317, App  |
| 27 | 19   | 63.3 | 843     | 13 | US-10-092-900A-119   | Sequence 119, App  |
| 28 | 19   | 63.3 | 40178   | 13 | US-10-282-174-467    | Sequence 467, App  |
| 29 | 19   | 63.3 | 40178   | 13 | US-10-282-174-468    | Sequence 468, App  |
| 30 | 19   | 63.3 | 495269  | 16 | US-10-398-221-8      | Sequence 8, Appl   |
| 31 | 19   | 63.3 | 3011208 | 16 | US-10-398-221-2058   | Sequence 2058, Ap  |
| 32 | 18.8 | 62.7 | 343     | 13 | US-10-027-632-177235 | Sequence 177235,   |
| 33 | 18.8 | 62.7 | 343     | 16 | US-10-027-632-177235 | Sequence 177235,   |
| 34 | 18.8 | 62.7 | 451     | 15 | US-10-198-846-4908   | Sequence 4908, Ap  |
| 35 | 18.8 | 62.7 | 574     | 15 | US-10-198-846-10080  | Sequence 10080, A  |
| 36 | 18.8 | 62.7 | 636     | 15 | US-10-198-846-4977   | Sequence 4977, Ap  |
| 37 | 18.8 | 62.7 | 646     | 15 | US-10-198-846-9104   | Sequence 9104, Ap  |
| 38 | 18.8 | 62.7 | 663     | 13 | US-10-027-632-231136 | Sequence 231136,   |
| 39 | 18.8 | 62.7 | 663     | 16 | US-10-027-632-231136 | Sequence 231136,   |
| 40 | 18.8 | 62.7 | 865     | 15 | US-10-198-846-4870   | Sequence 4870, Ap  |
| 41 | 18.8 | 62.7 | 868     | 15 | US-10-198-846-5195   | Sequence 5195, Ap  |
| 42 | 18.8 | 62.7 | 1143    | 9  | US-09-006-398-24     | Sequence 24, Appl  |
| 43 | 18.8 | 62.7 | 1440    | 14 | US-10-044-090-820    | Sequence 820, App  |
| 44 | 18.8 | 62.7 | 2331    | 16 | US-10-369-493-35883  | Sequence 35883, A  |
| 45 | 18.8 | 62.7 | 2907    | 9  | US-09-954-456-318    | Sequence 318, App  |
| 46 | 18.8 | 62.7 | 2907    | 9  | US-09-954-456-823    | Sequence 823, App  |
| 47 | 18.8 | 62.7 | 2907    | 9  | US-09-954-456-1226   | Sequence 1226, Ap  |
| 48 | 18.8 | 62.7 | 2907    | 9  | US-09-880-107-2318   | Sequence 2318, Ap  |
| 49 | 18.8 | 62.7 | 2907    | 13 | US-10-342-887-399    | Sequence 399, App  |
| 50 | 18.8 | 62.7 | 2907    | 13 | US-10-172-118-399    | Sequence 399, App  |
| 51 | 18.8 | 62.7 | 2907    | 17 | US-10-641-643-1033   | Sequence 1053, Ap  |
| 52 | 18.6 | 62.0 | 584     | 13 | US-10-027-632-253998 | Sequence 253998,   |
| 53 | 18.6 | 62.0 | 584     | 16 | US-10-027-632-253998 | Sequence 253998,   |
| 54 | 18.6 | 62.0 | 996     | 13 | US-10-440-503-219    | Sequence 219, App  |
| 55 | 18.6 | 62.0 | 996     | 15 | US-10-461-925-219    | Sequence 219, App  |
| 56 | 18.6 | 62.0 | 996     | 15 | US-10-146-772-219    | Sequence 219, App  |
| 57 | 18.6 | 62.0 | 996     | 16 | US-10-241-742-219    | Sequence 219, App  |
| 58 | 18.6 | 62.0 | 996     | 16 | US-10-440-523-219    | Sequence 219, App  |
| 59 | 18.6 | 62.0 | 1698    | 13 | US-10-425-114-1399   | Sequence 1399, Ap  |
| 60 | 18.6 | 62.0 | 1701    | 9  | US-09-801-368-417    | Sequence 417, App  |
| 61 | 18.6 | 62.0 | 2369    | 16 | US-10-108-260A-1248  | Sequence 1248, Ap  |
| 62 | 18.6 | 62.0 | 104083  | 13 | US-10-087-192-670    | Sequence 670, App  |
| 63 | 18.4 | 61.3 | 546     | 9  | US-09-919-580-732    | Sequence 732, App  |
| 64 | 18.4 | 61.3 | 634     | 13 | US-10-424-599-75105  | Sequence 75105, A  |
| 65 | 18.4 | 61.3 | 729     | 9  | US-09-764-855-160    | Sequence 160, App  |
| 66 | 18.4 | 61.3 | 729     | 15 | US-10-072-349-160    | Sequence 160, App  |
| 67 | 18.4 | 61.3 | 742     | 13 | US-10-027-632-141364 | Sequence 141364,   |
| 68 | 18.4 | 61.3 | 742     | 16 | US-10-027-632-141364 | Sequence 141364,   |
| 69 | 18.4 | 61.3 | 752     | 16 | US-10-264-049-2070   | Sequence 2070, Ap  |
| 70 | 18.4 | 61.3 | 778     | 15 | US-10-106-698-589    | Sequence 589, App  |
| 71 | 18.4 | 61.3 | 813     | 13 | US-10-107-632-160048 | Sequence 160048,   |
| 72 | 18.4 | 61.3 | 813     | 16 | US-10-027-632-160048 | Sequence 160048,   |
| 73 | 18.4 | 61.3 | 864     | 9  | US-09-070-927A-767   | Sequence 767, App  |
| 74 | 18.4 | 61.3 | 888     | 9  | US-09-764-855-159    | Sequence 159, App  |
| 75 | 18.4 | 61.3 | 888     | 15 | US-10-072-349-159    | Sequence 159, App  |
| 76 | 18.4 | 61.3 | 1059    | 11 | US-09-938-842A-1816  | Sequence 1816, Ap  |
| 77 | 18.4 | 61.3 | 1059    | 11 | US-09-938-842A-1816  | Sequence 1816, Ap  |
| 78 | 18.4 | 61.3 | 1246    | 10 | US-09-934-455-345    | Sequence 345, App  |
| 79 | 18.4 | 61.3 | 1246    | 13 | US-10-412-699B-1713  | Sequence 1713, App |
| 80 | 18.4 | 61.3 | 1246    | 13 | US-10-412-699B-1713  | Sequence 1713, App |
| 81 | 18.4 | 61.3 | 1246    | 13 | US-10-225-066A-587   | Sequence 587, App  |
| 82 | 18.4 | 61.3 | 1246    | 15 | US-10-278-173-123    | Sequence 123, App  |
| 83 | 18.4 | 61.3 | 1246    | 15 | US-10-295-403-7      | Sequence 7, Appl   |
| 84 | 18.4 | 61.3 | 1246    | 15 | US-10-278-536-171    | Sequence 171, App  |
| 85 | 18.4 | 61.3 | 1246    | 16 | US-10-374-780A-1949  | Sequence 1949, Ap  |
| 86 | 18.4 | 61.3 | 1246    | 16 | US-10-685-522-3      | Sequence 3, Appl   |
| 87 | 18.4 | 61.3 | 1315    | 16 | US-10-310-154-235    | Sequence 235, App  |

|       |      |      |         |    |                      |                    |       |      |      |         |    |                      |                   |
|-------|------|------|---------|----|----------------------|--------------------|-------|------|------|---------|----|----------------------|-------------------|
| C 88  | 18.4 | 61.3 | 2471    | 16 | US-10-367-980A-4     | Sequence 4, Appli  | C 161 | 17.8 | 59.3 | 866     | 15 | US-10-198-846-4450   | Sequence 4450, Ap |
| C 89  | 18.4 | 61.3 | 2855    | 10 | US-09-822-846-155    | Sequence 155, App  | C 162 | 17.8 | 59.3 | 946     | 12 | US-09-876-143-1726   | Sequence 1726, Ap |
| C 90  | 18.4 | 61.3 | 7058    | 15 | US-10-287-218-38     | Sequence 38, Appl  | C 163 | 17.8 | 59.3 | 1286    | 13 | US-10-027-632-230147 | Sequence 230147,  |
| C 91  | 18.4 | 61.3 | 7433    | 15 | US-10-198-846-13220  | Sequence 13220, A  | C 164 | 17.8 | 59.3 | 1286    | 16 | US-10-027-632-230147 | Sequence 230147,  |
| C 92  | 18.2 | 60.7 | 848     | 15 | US-10-029-386-24158  | Sequence 24158, A  | C 165 | 17.8 | 59.3 | 1299    | 9  | US-09-738-626-2305   | Sequence 2305, Ap |
| C 93  | 18.2 | 60.7 | 7931    | 16 | US-10-367-978-31     | Sequence 31, Appl  | C 166 | 17.8 | 59.3 | 1886    | 13 | US-10-020-120-31     | Sequence 31, Appl |
| C 94  | 18.2 | 60.7 | 14955   | 16 | US-10-144-198-29     | Sequence 29, Appl  | C 167 | 17.8 | 59.3 | 2000    | 16 | US-10-260-238-2369   | Sequence 2369, Ap |
| C 95  | 18   | 60.0 | 239     | 13 | US-10-424-599-1079   | Sequence 1079, Ap  | C 168 | 17.8 | 59.3 | 2303    | 15 | US-10-380-558-5      | Sequence 5, Appli |
| C 96  | 18   | 60.0 | 406     | 13 | US-09-878-722-124    | Sequence 124, App  | C 169 | 17.8 | 59.3 | 2477    | 13 | US-10-027-632-110037 | Sequence 110037,  |
| C 97  | 18   | 60.0 | 406     | 13 | US-09-878-722-124    | Sequence 124, App  | C 170 | 17.8 | 59.3 | 2477    | 13 | US-10-027-632-110038 | Sequence 110038,  |
| C 98  | 18   | 60.0 | 413     | 9  | US-09-960-352-7890   | Sequence 7890, App | C 171 | 17.8 | 59.3 | 2477    | 13 | US-10-027-632-110039 | Sequence 110039,  |
| C 99  | 18   | 60.0 | 424     | 9  | US-09-960-352-308    | Sequence 308, App  | C 172 | 17.8 | 59.3 | 2477    | 16 | US-10-027-632-110037 | Sequence 110037,  |
| C 100 | 18   | 60.0 | 427     | 9  | US-09-864-761-89     | Sequence 89, Appl  | C 173 | 17.8 | 59.3 | 2477    | 16 | US-10-027-632-110038 | Sequence 110038,  |
| C 101 | 18   | 60.0 | 468     | 9  | US-09-815-242-7062   | Sequence 7062, Ap  | C 174 | 17.8 | 59.3 | 2477    | 16 | US-10-027-632-110039 | Sequence 110039,  |
| C 102 | 18   | 60.0 | 468     | 13 | US-10-282-122A-22184 | Sequence 22184, A  | C 175 | 17.8 | 59.3 | 3125    | 14 | US-10-044-090-16     | Sequence 16, Appl |
| C 103 | 18   | 60.0 | 635     | 13 | US-10-027-632-14842  | Sequence 14842, A  | C 176 | 17.8 | 59.3 | 113604  | 13 | US-10-027-152B-1     | Sequence 1, Appli |
| C 104 | 18   | 60.0 | 635     | 16 | US-10-027-632-14842  | Sequence 14842, A  | C 177 | 17.8 | 59.3 | 113604  | 13 | US-10-027-152B-2     | Sequence 2, Appli |
| C 105 | 18   | 60.0 | 672     | 9  | US-09-864-761-16927  | Sequence 16927, A  | C 178 | 17.8 | 59.3 | 113604  | 15 | US-10-027-195A-1     | Sequence 1, Appli |
| C 106 | 18   | 60.0 | 824     | 9  | US-09-770-445-716    | Sequence 716, App  | C 179 | 17.8 | 59.3 | 113604  | 15 | US-10-027-195A-2     | Sequence 2, Appli |
| C 107 | 18   | 60.0 | 1062    | 9  | US-09-938-842A-2094  | Sequence 2094, Ap  | C 180 | 17.8 | 59.3 | 251364  | 15 | US-10-175-523-58     | Sequence 58, Appl |
| C 108 | 18   | 60.0 | 1062    | 11 | US-09-938-842A-2094  | Sequence 2094, Ap  | C 181 | 17.8 | 59.3 | 251364  | 15 | US-10-175-523-58     | Sequence 58, Appl |
| C 109 | 18   | 60.0 | 1136    | 9  | US-09-917-800A-1671  | Sequence 1671, Ap  | C 182 | 17.8 | 59.3 | 251364  | 15 | US-10-175-523-79     | Sequence 79, Appl |
| C 110 | 18   | 60.0 | 1136    | 12 | US-10-152-218A-1722  | Sequence 1722, Ap  | C 183 | 17.8 | 59.3 | 3309400 | 9  | US-09-738-626-1      | Sequence 1, Appli |
| C 111 | 18   | 60.0 | 1136    | 16 | US-10-191-803-121    | Sequence 121, App  | C 184 | 17.6 | 58.7 | 359     | 9  | US-09-728-445-651    | Sequence 651, App |
| C 112 | 18   | 60.0 | 1267    | 9  | US-09-770-445-25     | Sequence 25, Appl  | C 185 | 17.6 | 58.7 | 393     | 13 | US-10-424-599-89079  | Sequence 89079, A |
| C 113 | 18   | 60.0 | 1332    | 15 | US-10-022-390-660    | Sequence 660, App  | C 186 | 17.6 | 58.7 | 400     | 9  | US-09-789-919-18     | Sequence 18, Appl |
| C 114 | 18   | 60.0 | 1332    | 15 | US-10-022-390-667    | Sequence 667, App  | C 187 | 17.6 | 58.7 | 410     | 10 | US-09-803-719-1803   | Sequence 1803, Ap |
| C 115 | 18   | 60.0 | 1932    | 15 | US-10-022-390-688    | Sequence 688, App  | C 188 | 17.6 | 58.7 | 418     | 13 | US-10-424-599-88127  | Sequence 88127, A |
| C 116 | 18   | 60.0 | 1932    | 15 | US-10-022-390-691    | Sequence 691, App  | C 189 | 17.6 | 58.7 | 503     | 16 | US-10-341-961A-84    | Sequence 84, Appl |
| C 117 | 18   | 60.0 | 1932    | 15 | US-10-022-390-693    | Sequence 693, App  | C 190 | 17.6 | 58.7 | 514     | 9  | US-09-731-872-153    | Sequence 153, App |
| C 118 | 18   | 60.0 | 1932    | 15 | US-10-022-390-700    | Sequence 700, App  | C 191 | 17.6 | 58.7 | 514     | 10 | US-09-876-997-153    | Sequence 153, App |
| C 119 | 18   | 60.0 | 1932    | 15 | US-10-022-390-702    | Sequence 702, App  | C 192 | 17.6 | 58.7 | 588     | 15 | US-10-029-386-2780   | Sequence 2780, Ap |
| C 120 | 18   | 60.0 | 1932    | 15 | US-10-022-249-660    | Sequence 660, App  | C 193 | 17.6 | 58.7 | 631     | 9  | US-09-764-870-130    | Sequence 130, App |
| C 121 | 18   | 60.0 | 1932    | 15 | US-10-022-249-667    | Sequence 667, App  | C 194 | 17.6 | 58.7 | 631     | 11 | US-09-764-875-346    | Sequence 346, App |
| C 122 | 18   | 60.0 | 1932    | 15 | US-10-022-249-688    | Sequence 688, App  | C 195 | 17.6 | 58.7 | 631     | 15 | US-10-125-540-130    | Sequence 130, App |
| C 123 | 18   | 60.0 | 1932    | 15 | US-10-022-249-691    | Sequence 691, App  | C 196 | 17.6 | 58.7 | 654     | 15 | US-10-290-579-99     | Sequence 99, Appl |
| C 124 | 18   | 60.0 | 1932    | 15 | US-10-022-249-693    | Sequence 693, App  | C 197 | 17.6 | 58.7 | 685     | 13 | US-10-424-599-6185   | Sequence 6185, Ap |
| C 125 | 18   | 60.0 | 1932    | 15 | US-10-022-249-700    | Sequence 700, App  | C 198 | 17.6 | 58.7 | 929     | 17 | US-10-240-240A-26    | Sequence 26, Appl |
| C 126 | 18   | 60.0 | 1932    | 15 | US-10-022-249-702    | Sequence 702, App  | C 199 | 17.6 | 58.7 | 1448    | 15 | US-10-197-666A-15    | Sequence 15, Appl |
| C 127 | 18   | 60.0 | 2166    | 15 | US-10-101-510-326    | Sequence 326, App  | C 200 | 17.6 | 58.7 | 1448    | 15 | US-10-024-298A-38    | Sequence 38, Appl |
| C 128 | 18   | 60.0 | 4055    | 15 | US-10-233-045-9      | Sequence 9, Appli  | C 201 | 17.6 | 58.7 | 1448    | 15 | US-10-042-211A-38    | Sequence 38, Appl |
| C 129 | 18   | 60.0 | 4716    | 15 | US-10-084-817-237    | Sequence 237, App  | C 202 | 17.6 | 58.7 | 1448    | 17 | US-10-617-217A-38    | Sequence 38, Appl |
| C 130 | 18   | 60.0 | 87687   | 13 | US-10-087-192-1978   | Sequence 1978, Ap  | C 203 | 17.6 | 58.7 | 1622    | 15 | US-10-029-386-22666  | Sequence 22666, A |
| C 131 | 18   | 60.0 | 189998  | 16 | US-10-380-931-24     | Sequence 24, Appl  | C 204 | 17.6 | 58.7 | 1813    | 15 | US-10-037-270-29     | Sequence 29, Appl |
| C 132 | 18   | 60.0 | 197496  | 9  | US-09-877-177-10     | Sequence 10, Appl  | C 205 | 17.6 | 58.7 | 1813    | 16 | US-10-117-722-29     | Sequence 29, Appl |
| C 133 | 18   | 60.0 | 197496  | 13 | US-10-426-836-10     | Sequence 10, Appl  | C 206 | 17.6 | 58.7 | 2000    | 9  | US-09-887-576-837    | Sequence 837, App |
| C 134 | 18   | 60.0 | 493631  | 13 | US-10-087-192-205    | Sequence 205, App  | C 207 | 17.6 | 58.7 | 2000    | 16 | US-10-260-238-1733   | Sequence 1733, Ap |
| C 135 | 18   | 60.0 | 1830121 | 15 | US-10-329-960-1      | Sequence 1, Appli  | C 208 | 17.6 | 58.7 | 2167    | 9  | US-09-965-529-46     | Sequence 46, Appl |
| C 136 | 18   | 60.0 | 1830121 | 16 | US-10-329-960-1      | Sequence 1, Appli  | C 209 | 17.6 | 58.7 | 2167    | 10 | US-09-969-680A-46    | Sequence 46, Appl |
| C 137 | 17.8 | 59.3 | 285     | 9  | US-09-294-093B-2124  | Sequence 2124, Ap  | C 210 | 17.6 | 58.7 | 2343    | 16 | US-10-264-049-784    | Sequence 784, App |
| C 138 | 17.8 | 59.3 | 290     | 9  | US-09-294-093B-829   | Sequence 829, App  | C 211 | 17.6 | 58.7 | 2343    | 13 | US-10-282-122A-35133 | Sequence 35133, A |
| C 139 | 17.8 | 59.3 | 309     | 13 | US-10-424-599-96694  | Sequence 96694, A  | C 212 | 17.6 | 58.7 | 3530    | 9  | US-09-784-847-1985   | Sequence 1985, Ap |
| C 140 | 17.8 | 59.3 | 321     | 9  | US-09-960-352-6356   | Sequence 6356, Ap  | C 213 | 17.6 | 58.7 | 3530    | 15 | US-10-092-154-1495   | Sequence 1495, Ap |
| C 141 | 17.8 | 59.3 | 353     | 9  | US-09-960-352-451    | Sequence 451, App  | C 214 | 17.6 | 58.7 | 3530    | 13 | US-10-087-192-586    | Sequence 586, App |
| C 142 | 17.8 | 59.3 | 379     | 13 | US-10-425-114-25635  | Sequence 25635, A  | C 215 | 17.6 | 58.7 | 3536    | 15 | US-10-092-154-1984   | Sequence 1984, Ap |
| C 143 | 17.8 | 59.3 | 426     | 15 | US-10-198-846-4939   | Sequence 4939, Ap  | C 216 | 17.6 | 58.7 | 3536    | 15 | US-10-092-154-1492   | Sequence 1492, Ap |
| C 144 | 17.8 | 59.3 | 435     | 9  | US-09-880-107-2653   | Sequence 2653, App | C 217 | 17.6 | 58.7 | 5241    | 16 | US-10-260-238-1984   | Sequence 1984, Ap |
| C 145 | 17.8 | 59.3 | 450     | 15 | US-10-240-965-182    | Sequence 182, App  | C 218 | 17.6 | 58.7 | 5460    | 13 | US-10-282-122A-15354 | Sequence 15354, A |
| C 146 | 17.8 | 59.3 | 472     | 13 | US-10-085-783A-55343 | Sequence 55343, A  | C 219 | 17.6 | 58.7 | 9060    | 15 | US-10-290-579-10     | Sequence 10, Appl |
| C 147 | 17.8 | 59.3 | 472     | 16 | US-10-242-538A-55343 | Sequence 55343, A  | C 220 | 17.6 | 58.7 | 18825   | 9  | US-09-784-847-1495   | Sequence 1495, Ap |
| C 148 | 17.8 | 59.3 | 550     | 15 | US-10-198-846-12307  | Sequence 12307, A  | C 221 | 17.6 | 58.7 | 16825   | 13 | US-10-092-154-1495   | Sequence 1495, Ap |
| C 149 | 17.8 | 59.3 | 568     | 13 | US-10-027-633-22225  | Sequence 22225, A  | C 222 | 17.4 | 58.0 | 393     | 15 | US-10-029-386-24571  | Sequence 24571, A |
| C 150 | 17.8 | 59.3 | 568     | 13 | US-10-027-633-22225  | Sequence 22225, A  | C 223 | 17.4 | 58.0 | 455     | 13 | US-10-027-632-193089 | Sequence 193089,  |
| C 151 | 17.8 | 59.3 | 568     | 16 | US-10-027-633-22225  | Sequence 22225, A  | C 224 | 17.4 | 58.0 | 455     | 15 | US-10-027-632-193089 | Sequence 193089,  |
| C 152 | 17.8 | 59.3 | 568     | 16 | US-10-027-633-22225  | Sequence 22225, A  | C 225 | 17.4 | 58.0 | 455     | 16 | US-10-027-632-193089 | Sequence 193089,  |
| C 153 | 17.8 | 59.3 | 584     | 13 | US-10-027-633-186339 | Sequence 186339,   | C 226 | 17.4 | 58.0 | 483     | 13 | US-10-027-632-193089 | Sequence 193089,  |
| C 154 | 17.8 | 59.3 | 584     | 13 | US-10-027-633-186340 | Sequence 186340,   | C 227 | 17.4 | 58.0 | 483     | 13 | US-10-424-599-124170 | Sequence 124170,  |
| C 155 | 17.8 | 59.3 | 584     | 15 | US-10-221-945-8      | Sequence 8, Appli  | C 228 | 17.4 | 58.0 | 550     | 15 | US-10-029-386-10868  | Sequence 10868, A |
| C 156 | 17.8 | 59.3 | 584     | 16 | US-10-027-633-186339 | Sequence 186339,   | C 229 | 17.4 | 58.0 | 638     | 13 | US-10-424-599-5250   | Sequence 5250, Ap |
| C 157 | 17.8 | 59.3 | 584     | 16 | US-10-027-633-186340 | Sequence 186340,   | C 230 | 17.4 | 58.0 | 688     | 13 | US-10-425-114-26025  | Sequence 26025, A |
| C 158 | 17.8 | 59.3 | 625     | 13 | US-09-973-278-99348  | Sequence 99348, A  | C 231 | 17.4 | 58.0 | 712     | 13 | US-10-092-900A-117   | Sequence 117, App |
| C 159 | 17.8 | 59.3 | 663     | 13 | US-10-424-599-95348  | Sequence 95348, A  | C 232 | 17.4 | 58.0 | 776     | 13 | US-10-027-632-104125 | Sequence 104125,  |
| C 160 | 17.8 | 59.3 | 678     | 15 | US-10-198-846-1849   | Sequence 1849, Ap  | C 233 | 17.4 | 58.0 | 776     | 13 | US-10-027-632-325003 | Sequence 325003,  |



234 17.4 58.0 776 16 US-10-027-632-104125  
 235 17.4 58.0 776 16 US-10-027-632-325003  
 C 236 17.4 58.0 798 13 US-10-425-114-11559  
 C 237 17.4 58.0 904 15 US-10-138-846-3589  
 C 238 17.4 58.0 948 15 US-10-137-036-107  
 C 239 17.4 58.0 1314 10 US-09-968-433-48  
 C 240 17.4 58.0 1360 13 US-10-424-599-124168  
 C 241 17.4 58.0 1364 13 US-10-424-599-87503  
 C 242 17.4 58.0 1710 13 US-10-424-599-139445  
 C 243 17.4 58.0 1766 15 US-10-137-036-117  
 C 244 17.4 58.0 1981 13 US-10-425-114-2715  
 C 245 17.4 58.0 2008 13 US-10-425-114-21189  
 C 246 17.4 58.0 2068 16 US-10-104-047-63  
 C 247 17.4 58.0 2164 13 US-10-027-632-250839  
 C 248 17.4 58.0 2164 16 US-10-027-632-250839  
 C 249 17.4 58.0 2396 13 US-10-425-114-11406  
 C 250 17.4 58.0 2311 16 US-10-242-355-1174  
 C 251 17.4 58.0 2512 16 US-10-104-047-1888  
 C 252 17.4 58.0 2799 13 US-10-424-599-98143  
 C 253 17.4 58.0 3708 13 US-10-282-122A-24775  
 C 254 17.4 58.0 3939 13 US-10-170-385-416  
 C 255 17.4 58.0 3939 15 US-10-101-510-113  
 C 256 17.4 58.0 4464 9 US-09-738-626-1995  
 C 257 17.4 58.0 7158 9 US-09-974-308-2171  
 C 258 17.4 58.0 165361 16 US-10-085-117-241  
 C 259 17.4 58.0 194945 16 US-10-085-117-355  
 C 260 17.2 57.3 60 10 US-09-908-975-19254  
 C 261 17.2 57.3 214 9 US-09-864-761-30337  
 C 262 17.2 57.3 283 9 US-09-294-093B-5662  
 C 263 17.2 57.3 308 13 US-10-424-599-43021  
 C 264 17.2 57.3 315 13 US-10-424-599-20372  
 C 265 17.2 57.3 381 13 US-10-424-599-16271  
 C 266 17.2 57.3 404 16 US-10-027-632-45528  
 C 267 17.2 57.3 417 16 US-10-369-493-34763  
 C 268 17.2 57.3 419 13 US-10-027-632-309283  
 C 269 17.2 57.3 419 16 US-10-027-632-309283  
 C 270 17.2 57.3 438 15 US-10-198-846-12746  
 C 271 17.2 57.3 459 9 US-09-864-761-11125  
 C 272 17.2 57.3 470 13 US-10-027-632-35309  
 C 273 17.2 57.3 470 16 US-10-027-632-35309  
 C 274 17.2 57.3 488 10 US-09-918-995-26453  
 C 275 17.2 57.3 489 13 US-10-335-977-2911  
 C 276 17.2 57.3 500 10 US-09-918-995-1709  
 C 277 17.2 57.3 540 13 US-10-335-977-2912  
 C 278 17.2 57.3 546 13 US-10-027-632-5002  
 C 279 17.2 57.3 546 16 US-10-027-632-5002  
 C 280 17.2 57.3 552 13 US-10-424-599-111909  
 C 281 17.2 57.3 572 13 US-10-027-632-217089  
 C 282 17.2 57.3 572 16 US-10-027-632-217089  
 C 283 17.2 57.3 602 13 US-10-027-632-273182  
 C 284 17.2 57.3 602 16 US-10-027-632-273182  
 C 285 17.2 57.3 611 9 US-09-919-580-277  
 C 286 17.2 57.3 634 13 US-10-027-632-62359  
 C 287 17.2 57.3 634 13 US-10-027-632-297673  
 C 288 17.2 57.3 634 16 US-10-027-632-62359  
 C 289 17.2 57.3 634 16 US-10-027-632-297673  
 C 290 17.2 57.3 639 13 US-10-424-599-135447  
 C 291 17.2 57.3 807 13 US-10-027-632-165906  
 C 292 17.2 57.3 807 13 US-10-027-632-165906  
 C 293 17.2 57.3 807 13 US-10-027-632-165908  
 C 294 17.2 57.3 807 16 US-10-027-632-165908  
 C 295 17.2 57.3 807 16 US-10-027-632-165907  
 C 296 17.2 57.3 807 16 US-10-027-632-165908  
 C 297 17.2 57.3 846 16 US-10-291-265-660  
 C 298 17.2 57.3 902 16 US-10-291-265-188  
 C 299 17.2 57.3 903 16 US-10-369-493-43575  
 C 300 17.2 57.3 1041 13 US-10-027-632-248911

ALIGNMENTS

US-09-712-363-10  
 ; Sequence 10, Application US/09712363  
 ; Patent No. US20020164588A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Eisenberg, David  
 ; APPLICANT: Rotstein, Sergio H.  
 ; APPLICANT: Marcotte, Edward M.  
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 ; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
 ; FILE REFERENCE: 07419-032001  
 ; CURRENT APPLICATION NUMBER: US/09/712,363  
 ; CURRENT FILING DATE: 2000-11-13  
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246  
 ; PRIOR FILING DATE: 2000-01-28  
 ; PRIOR APPLICATION NUMBER: 60/179,531  
 ; PRIOR FILING DATE: 2000-02-01  
 ; PRIOR APPLICATION NUMBER: 60/117,844  
 ; PRIOR FILING DATE: 1999-01-29  
 ; PRIOR APPLICATION NUMBER: 60/118,206,  
 ; PRIOR FILING DATE: 1999-02-01  
 ; PRIOR APPLICATION NUMBER: 60/136,593  
 ; PRIOR FILING DATE: 1999-03-26  
 ; PRIOR APPLICATION NUMBER: 60/134,093  
 ; PRIOR FILING DATE: 1999-05-14  
 ; PRIOR APPLICATION NUMBER: 60/134,092  
 ; PRIOR FILING DATE: 1999-05-14  
 ; PRIOR APPLICATION NUMBER: 60/165,124  
 ; PRIOR FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: 60/165,086  
 ; PRIOR FILING DATE: 1999-11-12  
 ; NUMBER OF SEQ ID NOS: 292  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 10  
 ; LENGTH: 1584  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; US-09-712-363-10

Query Match 100.0%; Score 30; DB 9; Length 1584;  
 Best Local Similarity 100.0%; Pred. No. 0.0053;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 TGGCGCAGACCCAGGAGTAGCCCAATGAG 30  
 Db 441 TGGCGCAGACCCAGGAGTAGCCCAATGAG 470  
 RESULT 2  
 US-10-029-386-25513  
 ; Sequence 25513, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AECMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 25513  
 ; LENGTH: 176  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO CHR19.3  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4  
 ; OTHER INFORMATION: NT HIT: gi14572198, EVALUATE 1.00e-94  
 ; OTHER INFORMATION: EST\_HUMAN HIT: AW206860.1, EVALUATE 2.00e-63



```

; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; PRIOR FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 525
; LENGTH: 3050
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-525

Query Match      68.7%; Score 20.6; DB 13; Length 3050;
Best Local Similarity 85.2%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCAAT 27
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48

RESULT 9
US-10-172-118-525/c
; Sequence 525, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 525
; LENGTH: 3050
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_000767
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-525

Query Match      68.7%; Score 20.6; DB 13; Length 3050;
Best Local Similarity 85.2%; Pred. No. 61;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGGCGCAGAACCCAGGAGTAGCACCAAT 27
Db 74 TGGCGCTGAACCCAGGAGTAGCAAGAGT 48

RESULT 10
US-10-268-822-12/c
; Sequence 12, Application US/10268822
; Publication No. US20030150004A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Moore, David
; APPLICANT: Wei, Ping
; APPLICANT: Chua, Steven
; TITLE OF INVENTION: Screening Systems and Methods for Identifying Modulators of Xenob
; FILE REFERENCE: P02729US2
; CURRENT APPLICATION NUMBER: US/10/268,822
; CURRENT FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US 01/29672
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 10/219,590
; PRIOR FILING DATE: 2002-08-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 42547
; TYPE: DNA
; ORGANISM: Human
; US-10-268-822-12

Query Match      68.7%; Score 20.6; DB 15; Length 42547;
Best Local Similarity 85.2%; Pred. No. 89;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1 TGGCGCAGACCCAGGAGTAGCACCACAT 27
Db      4183 TGGCGCTGAACCCAGGAGTAGCAGAGT 4157

RESULT 11
US-10-027-632-22664/c
; Sequence 22664, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US 60/218,006
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22664
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-22664

Query Match      65.3%; Score 19.6; DB 16; Length 629;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      5 GCAGAACCCAGGAGTAGCACCACATGAG 30
Db      224 GCAGAACCCAGGAGTAGAACCACCCAGGAG 199

RESULT 12
US-10-027-632-22664/c
; Sequence 22664, Application US/10085117
; Publication No. US2003023234A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 99934
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)....(99934)
; OTHER INFORMATION: n = any nucleotide
; US-10-085-117-73

Query Match      65.3%; Score 19.6; DB 16; Length 99934;
Best Local Similarity 84.6%; Pred. No. 2.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1 TGGCGCAGACCCAGGAGTAGCACCACCA 26
Db      8930 TGGCGCAGGAGGAGTGCTCCAA 8905
```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7559
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019276.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
US-09-864-761-7559

Query Match 54.7%; Score 19.4; DB 9; Length 537;
Best Local Similarity 79.3%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GGCGCAGAACCGAGGAGTAGCACCACATGAG 30
Db 49 GGCCGAGAACCGAGGAGGCTCCAGTGAG 77

RESULT 16
US-10-371-264-15
; Sequence 15, Application US/10371264
; Publication No. US20030232061A1
; GENERAL INFORMATION:
; APPLICANT: Haller, Aurelia
; APPLICANT: Tang, Roderick
; TITLE OF INVENTION: RECOMBINANT PARAINFLUENZA VIRUS
; TITLE OF INVENTION: EXPRESSION SYSTEMS AND VACCINES
; TITLE OF INVENTION: COMPRISING HETEROLOGOUS ANTIGENS
; TITLE OF INVENTION: DERIVED FROM METAPNEUMOVIRUS
; FILE REFERENCE: 7682-067-999
; CURRENT APPLICATION NUMBER: US/10/371,264
; CURRENT FILING DATE: 2003-02-21
; PRIOR APPLICATION NUMBER: 60/358,934
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 327
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: rhinotracheitis virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)...(1260)
; OTHER INFORMATION: Turkey rhinotracheitis virus (strain 6574)
; OTHER INFORMATION: attachment protein (G), complete cds
US-10-371-264-15

Query Match 64.7%; Score 19.4; DB 16; Length 1260;
Best Local Similarity 79.3%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TGCGCGAGAACCGAGGAGTAGCACCACATGA 29
Db 1189 TGGCGAGAACCGAGGAGGCTCCAGTGAG 1217

```

|                       |                |                   |          |             |
|-----------------------|----------------|-------------------|----------|-------------|
| Query Match           | 64.7%          | Score 19.4        | DB 16    | Length 1260 |
| Best Local Similarity | 79.3%          | Pred. No. 1.7e+02 |          |             |
| Matches 23            | Conservative 0 | Mismatches 6      | Indels 0 | Gaps 0      |

  

|    |      |                              |      |
|----|------|------------------------------|------|
| Qy | 1    | TGGCGCAGACACCGAGTAGCACCAATGA | 29   |
|    |      |                              |      |
| Db | 1189 | TGGCGCAGACACCGAGATGCATCTGTGA | 1217 |



US-10-628-088-15

Query Match 64.7%; Score 19.4; DB 17; Length 1260;  
Best Local Similarity 79.3%; Pred. No. 1.7e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGAGACAGGAGTAGCACCATTGA 29  
DB 1189 TGGCGAGACAGGAGTAGCATTCTGTGA 1217

RESULT 21

US-10-369-493-25221/c  
; Sequence 25221, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 25221  
; LENGTH: 1779  
; TYPE: DNA  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-25221

Query Match 64.7%; Score 19.4; DB 16; Length 1779;  
Best Local Similarity 79.3%; Pred. No. 1.8e+02;  
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TGGCGAGACAGGAGTAGCACCATTGA 29  
DB 276 TGGCGAGATCAGGACACCAACCAATGA 248

RESULT 22

US-10-425-114-15923  
; Sequence 15923, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 15923  
; LENGTH: 1342  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3060-109-H1\_FLI  
US-10-425-114-15923

Query Match 64.0%; Score 19.2; DB 13; Length 1342;  
Best Local Similarity 87.5%; Pred. No. 2.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGCGAGAACGAGGAGTAGCACCA 25  
DB 45 GGCGAGAACGAGGAGTAGCACCA 68

RESULT 23

US-10-369-493-46459/c  
; Sequence 46459, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 46459  
; LENGTH: 1950  
; TYPE: DNA  
; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-46459

Query Match 64.0%; Score 19.2; DB 16; Length 1950;  
Best Local Similarity 87.5%; Pred. No. 2.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGCGAGAACGAGGAGTAGCACCA 25  
DB 1913 GGCGAGAACGAGGAGTAGCACCA 1890

RESULT 24

US-10-158-844-36  
; Sequence 36, Application US/10158844  
; Publication No. US20040029118A1  
; GENERAL INFORMATION:  
; APPLICANT: Kunsch et al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-R  
; COMPUTER: Dell Latitude Pentium 3  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/158,844  
; FILING DATE: 03-Jun-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/961,527  
; FILING DATE: 1997-10-30  
; APPLICATION NUMBER: US 60/029,960  
; FILING DATE: 1996-10-31  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hyman, Mark J.  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB340P1D1  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:





```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (132)..(771)
US-10-092-900A-119

Query Match      63.3%; Score 19; DB 13; Length 843;
Best Local Similarity 81.5%; Pred. No. 2.4e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGCGCGAGAACCCAGGAGTAGCACCACAT 27
Db 490 TGCGCGAGAACCCAGGAGTAGCACCACAT 454

RESULT 28
US-10-282-174-467/c
; Sequence 467, Application US/10282174
; Publication No. US20030224380A1
; GENERAL INFORMATION:
; APPLICANT: Becker, Kenneth David
; APPLICANT: Velicelebi, Gonul
; APPLICANT: Elliott, Kathryn J.
; APPLICANT: Wang, Xin
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Bertram, Lars
; APPLICANT: Saunders, Aleister J.
; APPLICANT: Mullin, Kristina M.
; APPLICANT: Sampson, Andrew Johnson
; APPLICANT: Blacker, Deborah Lynne
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ON CHROMOSOME 10
; TITLE OF INVENTION: ASSOCIATED WITH ALZHEIMER'S DISEASE AND OTHER
; FILE REFERENCE: 37481-3308
; CURRENT APPLICATION NUMBER: US 60/339,525
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 60/338,010
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/336,929
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/338,363
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/337,052
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/368,919
; NUMBER OF SEQ ID NOS: 564
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 40178
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1197,1841,1852,2075,6063,6173,6194,7219,7820,8242,10114,
; LOCATION: 10606,10688,10729,11559,12031,14437,14729,21145,21329,21404,
; OTHER INFORMATION: N is any
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1307,1308,1309
; OTHER INFORMATION: deletion: ATC
; FEATURE:
; NAME/KEY: allele
; LOCATION: (28453)...(28465)
; OTHER INFORMATION: deletion: TCCGCGAGAGGCG
; NAME/KEY: allele
; LOCATION: 21429,22246,22354,22621,23802,25283,25969,29904,39834,40018
; OTHER INFORMATION: N is any
US-10-282-174-468

Query Match      63.3%; Score 19; DB 13; Length 40178;
Best Local Similarity 81.5%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGAGTAGCACCACATG 28
Db 10145 GGCGCAGAACCCAGGAGTAGCACCACATG 10119

RESULT 30
US-10-398-221-8/c
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
```

```
; APPLICANT: Elliott, Kathryn J.
; APPLICANT: Wang, Xin
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Bertram, Lars
; APPLICANT: Saunders, Aleister J.
; APPLICANT: Mullin, Kristina M.
; APPLICANT: Sampson, Andrew Johnson
; APPLICANT: Blacker, Deborah Lynne
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ON CHROMOSOME 10
; TITLE OF INVENTION: ASSOCIATED WITH ALZHEIMER'S DISEASE AND OTHER
; FILE REFERENCE: 37481-3308
; CURRENT APPLICATION NUMBER: US/10/282,174
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/339,525
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 60/338,010
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/336,929
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/338,363
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/337,052
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/368,919
; NUMBER OF SEQ ID NOS: 564
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 468
; LENGTH: 40178
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1197,1841,1852,2075,6063,6173,6194,7219,7820,8242,10114,
; LOCATION: 10606,10688,10729,11559,12031,14437,14729,21145,21329,21404,
; OTHER INFORMATION: N is any
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1307,1308,1309
; OTHER INFORMATION: deletion: ATC
; FEATURE:
; NAME/KEY: allele
; LOCATION: (28453)...(28465)
; OTHER INFORMATION: deletion: TCCGCGAGAGGCG
; NAME/KEY: allele
; LOCATION: 21429,22246,22354,22621,23802,25283,25969,29904,39834,40018
; OTHER INFORMATION: N is any
US-10-282-174-468

Query Match      63.3%; Score 19; DB 13; Length 40178;
Best Local Similarity 81.5%; Pred. No. 4.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GGCGCAGAACCCAGGAGTAGCACCACATG 28
Db 10145 GGCGCAGAACCCAGGAGTAGCACCACATG 10119

RESULT 30
US-10-398-221-8/c
; Sequence 8, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
```

Mon Jun 21 09:02:14 2004

```

; PRIOR APPLICATION NUMBER: PR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 495269
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-8

Query Match      63.3%; Score 19; DB 16; Length 495269;
Best Local Similarity 81.5%; Pred. No. 6e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      3  GCGCAGAACCGAGGAGTAGCACCACATGA 29
         |||||
Db      439229 GCGCAGAACCGAGGAGTAGCACCACATGA 439203

Search completed: June 20, 2004, 17:46:47
Job time : 178.683 secs

```